

И2
Т0
Н2
Т2



СБОРНИК ТРУДОВ

**VIII Международной конференции
и молодежной школы «Информационные
технологии и нанотехнологии»**



**23-27 МАЯ
САМАРА, РОССИЯ**

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

ИНСТИТУТ СИСТЕМ ОБРАБОТКИ ИЗОБРАЖЕНИЙ РАН –
ФИЛИАЛ ФНИЦ "КРИСТАЛЛОГРАФИЯ И ФОТОНИКА" РАН

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
И НАНОТЕХНОЛОГИИ (ИТНТ-2022)
Том 5. Науки о данных

*Сборник трудов по материалам
VIII Международной конференции и молодежной школы
(г. Самара, 23-27 мая)*

Одобрено редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева»

© Самарский университет, 2022

ISBN 978-5-7883-1788-5

ISBN 978-5-7883-1793-9 (т. 5)

САМАРА
Издательство Самарского университета
2022

УДК 004.9
ББК 32.973
И741

И741 Информационные технологии и нанотехнологии (ИТНТ-2022): сборник трудов по материалам VIII Международной конференции и молодежной школы (г. Самара, 23-27 мая): в 5 томах / Министерство науки и высшего образования Российской Федерации, Самарский университет, Институт систем обработки изображений РАН – филиал ФНИЦ "Кристаллография и фотоника" РАН. – Самара: Издательство Самарского университета, 2022. – **Том 5. Науки о данных** / под ред. А. В. Куприянова. – 1 CD-ROM (6,8 Мб). – Загл. с титул. экрана. – Текст. Изображение: электронный.

ISBN 978-5-7883-1793-9 (т. 5)
ISBN 978-5-7883-1788-5

Тематика Конференции ИТНТ-2022 охватывает широкий круг областей применения информационных технологий в науке и высокотехнологичных отраслях промышленности. Одним из приоритетных направлений работы Конференции является образовательный аспект, заключающийся в предоставлении студентам и молодым ученым возможности ознакомиться с новейшими научными достижениями по тематике Конференции, а также с уникальным научным оборудованием и лабораторной базой Самарского университета, используемой для реализации современных научных проектов.

УДК 004.9
ББК 32.973

Минимальные системные требования:

PC, процессор Pentium, 160 МГц; оперативная память 32 Мб;
на винчестере 16 Мб; Microsoft Windows XP/Vista/7;
разрешение экрана 1024x768 с глубиной цвета 16 бит;
DVD-ROM2-х и выше, мышь; Adobe Acrobat Reader.

Редактор тома А.В. Куприянов

Выпускающий редактор Д.В. Кудряшов

Подписано для тиражирования 13.07.2022.

Объем издания 6,8 Мб.

Количество носителей 1 диск.

Тираж 11 дисков.

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

443086, САМАРА, МОСКОВСКОЕ ШОССЕ, 34.

Издательство Самарского университета.
443086, Самара, Московское шоссе, 34.

Оглавление

Предисловие	8-11
1. Исследование нейрофизиологических закономерностей болезни Паркинсона на первой стадии методом анализа всплескообразной электрической активности О.С. Сушкова, А.А. Морозов, А.В. Габова, А.В. Карабанов, С.Н. Иллариошкин	050032
2. Визуальная среда разработки и визуальное программирование как эффективный инструмент для работы с данными В.Ф. Зинатуллин, С.Н. Коледин	050192
3. Моделирование и анализ данных конкурентоспособности территорий при переходе к экономике чистой энергии И.Н. Хаймович, В.М. Рамзаев, В.Г. Чумак	050203
4. Пространственно-временной скейлинг зрительно-вызванных нейромагнитных сигналов человека В.А. Юнусов, С.А. Демин, И.А. Русанова, А.В. Минкин, А.А. Еленёв	050232
5. Мультиклассовая классификация для диагностики функционирования сложных технических систем Ю.Е. Кувайскова, Н.А. Ломовцева, В.Н. Клячкин	050242
6. Использование машинного обучения для управления навигацией робота по показаниям датчиков расстояний Н.А. Ломовцева, Ю.Е. Кувайскова, И.Н. Карпунина, В.Н. Клячкин, Д.М. Ястребов	050252
7. Алгоритм выбора тарифов страхования промышленных рисков на основе анализа больших массивов данных об объектах страхования звание статьи М.И. Гераськин, Е.П. Ростова	050582
8. Разработка методики сборки соплового аппарата с использованием методов кластеризации В.А. Печенин, Е.Ю. Печенина, А.В. Куприянов	050602
9. Классификация по прецедентам и поиск в данных частых элементов Н.А. Драгунов, Е.В. Дюкова, А.П. Дюкова	050712
10. Метод комплексного анализа природных данных с применением искусственного интеллекта В.В. Гешпенер, Б.С. Мандрикова, Н.Г. Пэтайчук	050862
11. Анализ очередей для видеотрафика с помощью обобщённого интервального метода Б.Я. Лихтциндер, Е.В. Китаева, А.Ю. Привалов	050942
12. Оценка параметров движения объекта в условиях изменения качества измерительных данных Ю.В. Цыганова, А.А. Калянов, О.В. Лукин	050992
13. Параметрическая идентификация распределенной мультисенсорной системы фильтрации И.В. Семушин, Ю.В. Цыганова, А.В. Цыганов	051002
14. Исследование погрешности разностного решения уравнения теплопроводности в многослойной среде методом вычислительного эксперимента Ю.Ю. Кривошеева	051012
15. Agent-Oriented simulation modeling of systems М.А. Nikitina	051062
16. Разработка нагрузочного теста для исследования производительности распределенных файловых систем на основе подсчета ортогональных пар диагональных латинских квадратов С.Н. Попов, С.В. Востокин	051112

17. Выявление признаков объекта на основе неполных и противоречивых эмпирических данных В.А. Семенова, С.В. Смирнов	051122
18. Разработка информационной системы для повышения эффективности эксплуатации нефтяных скважин Г.В. Спиридонов, В.В. Мокшин, Э.Х. Шамсиев	051192
19. Разложение in vivo рамановских спектров кожи методом разрешения многомерных кривых И.А. Матвеева, О.О. Мякинин, И.А. Братченко	051252
20. Новые алгоритмы дискретной фильтрации на основе MWGS-ортогонализации для систем с мультипликативными и аддитивными шумами А.В. Цыганов, Ю.В. Цыганова, А.Н. Кувшинова, Т.Н. Куренева	051332
21. Разработка приложения по моделированию локомоторной активности лабораторных животных Д.С. Борисов, А.В. Благоев	051512
22. Оценка качества интерфейса на основе анализа временных рядов результатов обучения пользователей В.К. Григорьев, М.А. Овчинников	051562
23. Исследование и разработка алгоритма минимизации звуковой петли в системах громкоговорящей связи А.А. Колпаков, Д.В. Бейлекчи, А.Ю. Проскуряков, А.А. Белов	051662
24. Спектральный анализ стабильных графических сигналов методами Фурье и Гильберта – Хуанга М.В. Дакинова, Л.М. Бикчентаева, О.А. Саченков, Т.В. Балтина, Г.Г. Яфарова	051812
25. Метод оценки движения камеры на основе компланарности, инвариантный к глубине сцены Е.В. Гошин	051822
26. Сравнение алгоритмов отбора признаков для задач классификации данных М.Д. Тисленко, А.В. Гайдель	051833
27. Распределенная реализация метода средних решающих правил с умными выборками для больших задач SVM М.Ю. Курбаков, А.В. Копылов, С.Д. Двоенко, В.В. Сулимова	051842
28. Анализ социально-экономического развития региона на базе цифровой платформы интегрального мониторинга П.В. Ситников, Е.А. Додонова, И.Н. Дубинина, А.В. Иващенко, О.К. Головин, Н.Ю. Ильцова	052032
29. Применение технологий удаленных вычислений для исследования физико-химических процессов Э.Н. Мифтахов, С.А. Мустафина, Т.А. Михайлова, А.Х. Даминов	052062
30. Анализ локальных корреляционных характеристик биоэлектрических сигналов человека при выполнении когнитивных задач В.А. Юнусов, С.А. Демин	052102
31. Выбор стратегии поведения участника электронных торгов в сфере государственных закупок с помощью прогнозирования Ю.В. Хицкова, К.А. Маковий, И.Ф. Астахова, О.Е. Ефимова	052122
32. Двухкаскадные адаптивные подавители помех с управляемыми интервалами адаптации В.А. Засов, П.А. Мельников	052152
33. Решение задачи потокораспределения гидравлических сетей в условиях недостоверных исходных данных А.О. Хуснутдинов, В.С. Карманов	052192

34. Децентрализованный алгоритм обнаружения изменения режима движения объекта по данным мультисенсоров
А.В. Цыганов, Ю.В. Цыганова, А.В. Голубков 052212
35. Использование нейросетевой модели для исследования нефтяных объектов
А.А. Минигалиев, И.М. Кычкин, В.В. Мокшин 052272
36. Технология применения метода поиска информативных признаков для снижения размерности признакового пространства в задаче классификации областей натуральных гиперспектральных изображений
М.И. Хотилин 052362
37. Анализ алгоритмов поиска подстроки в тексте с учетом их практического применения на основе экспериментов
В.А. Михайлов, Т.А. Михайлова, Э.Н. Мифтахов, С.А. Мустафина 052392
38. Исследование важности входных признаков при прогнозировании геомагнитного индекса алгоритмами машинного обучения
Р.Д. Владимиров, В.Р. Широкий, О.Г. Баринов, И.Н. Мягкова, С.А. Доленко 052522
39. Разработка метода оценки качества средств технической защиты информации
Л.С. Элбакян, Л.С. Элбакян 052622
40. Определение вероятности потери информации на частично записанном оптическом диске однократной записи при длительном хранении в электронном архиве
А.В. Чернышов 052702
41. Применение цифровых двойников для анализа прочности костей на основе данных компьютерной томографии
О.В. Герасимов, К.Р. Шарафутдинова, Р.Р. Рахматулин, Т.В. Балтина, М.Э. Балтин, А.О. Федянин 052752
42. Повышение достоверности обнаружения состояний гонки данных многопоточных системах
А.Ю. Лобачев, В.А. Засов 052782
43. Анализ методов отбора признаков для задачи классификации когнитивных состояний и состояний покоя по данным ЭЭГ
М.Ш. Муртазина, Т.В. Авдеенко 052912
44. Статистический анализ временных рядов температуры воздуха приземного слоя атмосферы в различных регионах России
В.А. Газарян, П.И. Кудрявцев, А.В. Безрукова, Ю.А. Курбатова, Н.Е. Шапкина, А.И. Чуличков 052972
45. Алгоритм оценки продолжительности стадий заболевания по набору неполных данных
П.Н. Подзолков 053212
46. Программный инструментальный нейрокомпьютерного интерфейса для анализа электроэнцефалограмм на базе непрерывного вейвлет-преобразования
А.А. Столбова, В.Ю. Дьяконов, О.К. Головнин 053232
47. Прогнозирование состояния технического объекта на основе модели системы квазипериодических процессов в виде изображений на цилиндре
В.Р. Крашенинников, Ю.Е. Кувайскова, В.Н. Клячкин 053252
48. Алгоритмическое обеспечение персонального виртуального ассистента для автоматизации обработки клиентских запросов
К.С. Добратулин, М.И. Нежурина 053292

49. Scalogram-EMD distance for mobile ECGs
V. Guryanova 053352
50. Oblivious piecewise-linear decision trees
A. Gurianov 053462
51. Платформа имитационного моделирования оборудования инфраструктурных объектов
в распределенной среде
И.В. Бычков, А.Г. Феоктистов, Р.О. Костромин 053772

ПРЕДИСЛОВИЕ

Конференция ИТНТ-2022 проводится с целью предоставления возможности научных дискуссий и обсуждения результатов фундаментальных и прикладных исследований в области информационных технологий и нанотехнологий, привлечения молодежи в сферу передовых научных исследований, обмена опытом научно-образовательной деятельности при подготовке ИТНТ-специалистов.

Тематика Конференции ИТНТ-2022 охватывает широкий круг областей применения информационных технологий в науке и высокотехнологичных отраслях промышленности.

Основными направлениями работы Конференции ИТНТ-2022 являются:

Компьютерная оптика и нанофотоника

- дифракционная оптика;
- планарные оптические структуры;
- гиперспектральные системы;
- нанофотоника;
- системы оптической сенсорики, передачи и обработки информации;
- сингулярная оптика.

Информационные технологии дистанционного зондирования Земли

- информационные технологии в проектировании космических аппаратов дистанционного зондирования Земли и полезных нагрузок для них;
- программные и математические решения для управления движением космических аппаратов наблюдения;
- программные и аппаратные средства для получения, обработки и анализа данных, получаемых с космических аппаратов дистанционного зондирования Земли;
- математическое моделирование процессов функционирования космических аппаратов дистанционного зондирования Земли;
- современные проектные решения для создания космических аппаратов мониторинга Земли и околоземного пространства и их группировок, в том числе на базе аппаратов типа CubeSat;
- системы дистанционного зондирования Земли на основе БПЛА.

Распознавание, обработка и анализ изображений

- математические методы цифровой обработки изображений и распознавания образов
- трёхмерное зрение
- биометрические системы на основе изображений
- геоинформационные системы и технологии
- защита и верификация мультимедиа

Искусственный интеллект

- новые подходы, тренды и фундаментальные результаты в сфере искусственного интеллекта и его приложениях к распознаванию образов и анализу изображений, обработке текстов, речевой информации;
- нейросетевые методы и глубокое обучение;
- прикладные технологии искусственного интеллекта в обработке изображений, беспилотном транспорте, производственных и сельскохозяйственных приложениях, медицинских приложениях, экологии, мониторинге окружающей среды и других;
- программные технологии для решения задач искусственного интеллекта – фреймворки, библиотеки, открытые инициативы и сообщества;
- мультидисциплинарные аспекты искусственного интеллекта и машинного обучения.

Науки о данных

- Компьютерные науки:
 - инженерия данных;
 - визуализация данных;
 - математические методы анализа данных;
 - программные платформы и библиотеки для работы с данными;
 - аппаратные средства хранения и обработки данных;
 - высокопроизводительные, параллельные и облачные вычисления, технологии обработки больших данных;
 - базы данных, инструменты и языки для работы с базами данных.
- Прикладные задачи интеллектуального анализа данных:
 - решение актуальных прикладных задач.

Одним из приоритетных направлений работы Конференции ИТНТ-2022 является образовательный аспект, заключающийся в предоставлении студентам и молодым ученым возможности ознакомиться с новейшими научными

достижениями по [тематике](#) Конференции, а также с уникальным научным оборудованием и лабораторной базой [Самарского университета](#), используемой для реализации современных научных проектов.

В рамках Конференции проводится Молодежная школа, где молодые ученые и студенты получают возможность повысить свой профессиональный уровень и [опубликовать](#) свои научные результаты.

В данный сборник трудов вошли материалы по 5 основным направлениям Конференции:

Том 1. Компьютерная оптика и нанофотоника (под редакцией к.ф.-м.н. Е.С. Козловой)

Том 2. Информационные технологии дистанционного зондирования Земли (под редакцией к.т.н. И.С. Ткаченко)

Том 3. Распознавание, обработка и анализ изображений (под редакцией д.ф.-м.н. В.В. Мясникова)

Том 4. Искусственный интеллект (под редакцией д.т.н. А.В. Никонорова)

Том 5. Науки о данных (под редакцией д.т.н. А.В. Куприянова).

Выпускающий редактор томов 1-5: Д.В. Кудряшов.

Официальный сайт Конференции ИТНТ-2022: <http://itnt-conf.org/>

ОРГАНИЗАТОРЫ

- Институт систем обработки изображений РАН (ИСОИ РАН) – филиал ФНИЦ «Кристаллография и фотоника» РАН, г. Самара, Россия;
- Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет), г. Самара, Россия.

Организационный комитет

Председатель

Богатырёв В.Д. – д.э.н., профессор, ректор Самарского национального исследовательского университета, г. Самара, Россия

Заместители председателя

Казанский Н.Л. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;

Сергеев В.В. – д.т.н., профессор, Самарский университет, г. Самара, Россия;

Куприянов А.В. – д.т.н., профессор, Самарский университет, г. Самара, Россия.

Ответственный секретарь

Хабибуллин Р.М. – к.т.н., Самарский университет, г. Самара, Россия.

Члены Организационного комитета

Бланк В.А. – к.ф.-м.н., Самарский университет, г. Самара, Россия;

Бояркин Ю.Н. – ИСОИ РАН, г. Самара, Россия;

Востокин С.В. – д.т.н., Самарский университет, г. Самара, Россия;

Гашников М.В. – к.т.н., Самарский университет, г. Самара, Россия;

Гошин Е.В. – к.т.н., Самарский университет, г. Самара, Россия;

Жердев Д.А. – к.т.н., ИСОИ РАН, г. Самара, Россия;

Кадомина Е.А. – Самарский университет, г. Самара, Россия;

Кириш Д.В. – к.т.н., Самарский университет, г. Самара, Россия;

Котов А.П. – к.т.н., ИСОИ РАН, г. Самара, Россия;

Кузнецов А.В. – к.т.н., Самарский университет, г. Самара, Россия;

Куприянов Д.Д. – Самарский университет, г. Самара, Россия;

Максимов А.И. – Самарский университет, г. Самара, Россия;

Мисиевич С.К. – Самарский университет, г. Самара, Россия;

Паренский Н.А. – Самарский университет, г. Самара, Россия;

Подлипов В.В. – ИСОИ РАН, г. Самара, Россия;

Попов С.Б. – д.т.н., проф., Самарский университет, г. Самара, Россия;

Пресняков К.Г. – Департамент информационных технологий и связи Самарской области, г. Самара, Россия;

Рыцарев И.А. – Самарский университет, г. Самара, Россия;

Савельева А.А. – Самарский университет, г. Самара, Россия;

Скиданова Е.Б. – Самарский университет, г. Самара, Россия;

Смагин С.В. – ИСОИ РАН, г. Самара, Россия;

Стафеев С.С. – к.ф.-м.н., ИСОИ РАН, г. Самара, Россия;

Тиц С.Н. – к.т.н., Самарский университет, г. Самара, Россия;

Ткаченко И.С. – к.т.н., доцент, Самарский университет, г. Самара, Россия;

Хнырева Е.С. – Самарский университет, г. Самара, Россия;

Фомченков С.А. – Самарский университет, г. Самара, Россия;

Якуненкова Д.М. – ИСОИ РАН, г. Самара, Россия.

Программный комитет

Председатель

Сойфер В.А. – академик РАН, д.т.н., профессор, президент Самарского университета, Самара, Россия.

Заместитель председателя

Казанский Н.Л. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия.

Члены Программного комитета

Korotkova O. – Prof., University of Miami, Coral Gables, USA;
Niemann H. – Prof., Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany;
O'Faolain L. – Prof., Tyndall National Institute, Cork, Ireland;
Sazhin S. – Prof., University of Brighton, Brighton, United Kingdom;
Sobolewski M. – Prof., Polish-Japanese Institute of IT, Warsaw, Poland;
Калачёв Л.В. – профессор, Университет Монтаны, г. Миссула, США;
Коробейников А. – профессор, Педагогический университет Шеньси, г. Сиань, Китай;
Бычков И.В. – академик РАН, д.т.н., профессор, Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения РАН, г. Иркутск, Россия;
Воеводин В.В. – член-корреспондент РАН, д.ф.-м.н., профессор, Московский государственный университет имени М.В. Ломоносова, г. Москва, Россия;
Головашкин Д.Л. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;
Гуляев Ю.В. – академик РАН, д.ф.-м.н., профессор, Институт радиотехники и электроники им. В.А. Котельникова РАН, г. Москва, Россия;
Желтов С.Ю. – академик РАН, д.т.н., профессор, ГосНИИ авиационных систем, г. Москва, Россия;
Журавлев Ю.И. – академик РАН, д.ф.-м.н., профессор, Вычислительный центр им. А.А. Дородницына РАН, г. Москва, Россия;
Калошин В.А. – д.ф.-м.н., профессор, Институт радиотехники и электроники им. В.А. Котельникова РАН, г. Москва, Россия;
Козлова Е.С. – к.ф.-м.н., ИСОИ РАН, г. Самара, Россия.
Конов В.И. – академик РАН, д.ф.-м.н., профессор, Институт общей физики им. А.М. Прохорова РАН, г. Москва, Россия;
Котляр В.В. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;
Кульчин Ю.Н. – академик РАН, д.ф.-м.н., профессор, Институт автоматизации и процессов управления Дальневосточного отделения РАН, г. Владивосток, Россия;
Куприянов А.В. – д.т.н., профессор, Самарский университет, г. Самара, Россия;
Лабунец В.Г. – д.т.н., профессор, Уральский государственный лесотехнический университет, г. Екатеринбург, Россия;
Мясников В.В. – д.ф.-м.н., профессор, Самарский университет, г. Самара, Россия;
Никитов С.А. – член-корреспондент РАН, д.ф.-м.н., профессор, Институт радиотехники и электроники им. В.А. Котельникова РАН, г. Москва, Россия;
Никоноров А.В. – д.т.н., ИСОИ РАН, г. Самара, Россия;
Новиков Д.А. – чл.-корр. РАН, д.т.н., профессор, Института проблем управления РАН, г. Москва, Россия;
Потатуркин О.И. – д.т.н., профессор, Институт автоматизации и электрометрии Сибирского отделения Российской академии наук, г. Новосибирск, Россия;
Ткаченко И.С. – к.т.н., доцент, Самарский университет, г. Самара, Россия;
Хонина С.Н. – д.ф.-м.н., профессор, ИСОИ РАН, г. Самара, Россия;
Чочиа П.А. – д.т.н., Институт проблем передачи информации им. А.А. Харкевича РАН, г. Москва, Россия.

Исследование нейрофизиологических закономерностей болезни Паркинсона на первой стадии методом анализа всплескообразной электрической активности

О.С. Сушкова
Институт радиотехники и
электроники им. В.А. Котельникова
РАН
Москва, Россия
o.sushkova@mail.ru

А.А. Морозов
Институт радиотехники и
электроники им. В.А. Котельникова
РАН
Москва, Россия
morozov@cplire.ru

А.В. Габова
Институт высшей нервной
деятельности и нейрофизиологии
РАН
Москва, Россия
agabova@yandex.ru

А.В. Карабанов
Научный центр неврологии
Москва, Россия
doctor.karabanov@mail.ru

С.Н. Иллариошкин
Научный центр неврологии
Москва, Россия
snillario@gmail.com

Аннотация—Метод анализа всплескообразной электрической активности применён для изучения нейрофизиологических закономерностей в электромиографических сигналах у пациентов с болезнью Паркинсона на первой стадии. Исследованы параметры разных видов тремора. Обнаружена статистически значимая отрицательная корреляция между количеством всплесков паркинсонического и физиологического тремора у пациентов на первой стадии болезни Паркинсона, а также статистически значимая отрицательная корреляция между количеством всплесков физиологического тремора и возрастом пациентов.

Ключевые слова— болезнь Паркинсона, электромиограмма, ЭМГ, вейвлет-спектрограмма, метод анализа всплескообразной электрической активности, тремор конечностей, AUC-диаграммы, ROC-анализ.

1. ВВЕДЕНИЕ

Для исследования нейрофизиологических закономерностей в электромиографических (ЭМГ) сигналах у пациентов с болезнью Паркинсона (БП) был применён инструментальный разработанный ранее авторами метода анализа всплескообразной электрической активности. Изначально параметры всплескообразной электрической активности в сигналах изучались авторами в качестве метрик, позволяющих эффективно распознавать некоторые нейродегенеративные заболевания, такие как БП и эссенциальный тремор (ЭТ). В настоящем исследовании нами показано, что данные метрики отражают нейрофизиологические закономерности протекания БП на первой стадии. Были исследованы параметры разных видов тремора, а именно, паркинсонического (3-7 Гц) и физиологического (8-20 Гц) тремора. С помощью анализа всплескообразной электрической активности была обнаружена отрицательная корреляция между количеством всплесков паркинсонического и физиологического тремора у пациентов на первой стадии БП. Кроме того, была обнаружена корреляция между количеством всплесков физиологического тремора и возрастом пациентов.

2. МАТЕРИАЛЫ И МЕТОДЫ

Осуществлялось сравнение групп пациентов на первой стадии БП с тремором левой руки (10 человек) и тремором правой руки (12 человек) с группой здоровых добровольцев (10 человек).

Для анализа параметров всплесков использованы AUC-диаграммы [1-7]. Ранее эта техника была успешно применена для анализа электроэнцефалографических сигналов [2-4] и сигналов акселерометра [7]. На Рис. 1 и 2 приведены результаты анализа.

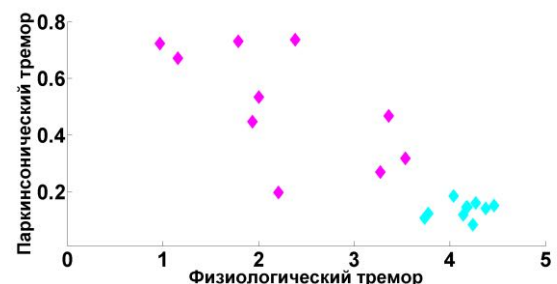


Рис. 1. Диаграммы рассеяния количества всплесков для левых рук пациентов с БП

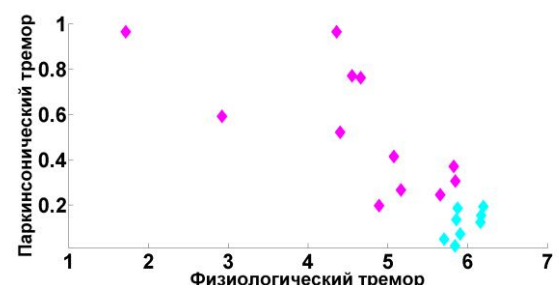


Рис. 2. Диаграммы рассеяния количества всплесков для правых рук пациентов с БП

Диаграммы рассеяния демонстрируют количество всплесков в секунду, обнаруженных в мышце-разгибателе треморных рук пациентов с БП. По оси

абсцисс отложено количество всплесков в секунду, соответствующих частотному диапазону физиологического тремора; по оси ординат – паркинсонического тремора. Пациенты с БП обозначены фиолетовым цветом. Контрольные испытуемые обозначены голубым цветом. На Рис. 1 показана диаграмма для левой руки испытуемых. На Рис. 2 – для правой руки испытуемых. Облако точек пациентов с БП расположено перпендикулярно диагонали диаграммы рассеяния, что свидетельствует об отрицательной корреляции между количеством всплесков, соответствующих частотам физиологического и паркинсонического тремора. Корреляция является статистически значимой у пациентов с тремором правой руки (Рис. 2); коэффициент корреляции Спирмена -0.75 , $p\text{-value} < 0.01$. Исследование всплесков у пациентов с БП с тремором левой руки выявило статистическую тенденцию (коэффициент корреляции Спирмена -0.5 , $p\text{-value} < 0.16$) (Рис. 1). Отрицательная корреляция между количеством всплесков в частотных диапазонах, соответствующих паркинсоническому и физиологическому тремору, указывает на взаимное отрицательное влияние некоторых нейрофизиологических механизмов, лежащих в основе обоих типов тремора.

Также отметим, что анализ корреляции между количеством всплесков в частотном диапазоне физиологического тремора на мышцах-разгибателях и возрастом пациентов с БП выявил существенные различия между пациентами с правосторонним и левосторонним дебютом БП. Диаграмма рассеяния (Рис. 3) демонстрирует связь между количеством всплесков в левых треморных руках и возрастом пациентов с БП. По оси абсцисс – возраст пациентов; по оси ординат – количество всплесков в секунду, соответствующих частотному диапазону физиологического тремора. Статистически значимая корреляция была обнаружена только у пациентов с БП с тремором левой руки. Коэффициент корреляции Спирмена составил -0.74 , $p\text{-value} < 0.02$.

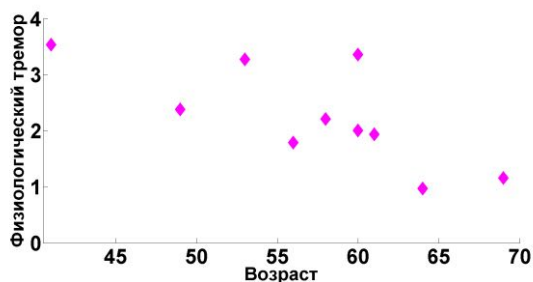


Рис. 3. Зависимость количества всплесков от возраста пациентов с БП. Левые треморные руки

3. ЗАКЛЮЧЕНИЕ

Обнаружены новые нейрофизиологические закономерности в данных ЭМГ у пациентов с БП. Выявленные закономерности проливают свет на нейрофизиологические процессы, протекающие в коре головного мозга, и позволяют выдвинуть гипотезы [1], объясняющие механизмы взаимодействия различных структур мозга у пациентов с БП.

БЛАГОДАРНОСТИ

Авторы выражают благодарность д.ф.-м.н. Обухову Ю.В. за помощь в проведении исследований. Работа выполнена в рамках государственного задания.

ЛИТЕРАТУРА

- [1] Sushkova, O.S. A statistical method for exploratory data analysis based on 2D and 3D area under curve diagrams: Parkinson's disease investigation / O.S. Sushkova, A.A. Morozov, A.V. Gabova, A.V. Karabanov, S.N. Illarionov // *Sensors*. – 2021. – Vol. 21(14). – P. 4700. DOI: 10.3390/s21144700.
- [2] Sushkova, O.S. Data mining in EEG wave trains in early stages of Parkinson's disease / O.S. Sushkova, A.A. Morozov, A.V. Gabova // *Advances in Soft Computing: Proceedings 15th Mexican International Conference on Artificial Intelligence, MICAI, Cancun, Mexico*. – Cham: Springer International Publishing, 2017. – Vol. 2. – P. 403-412. DOI: 10.1007/978-3-319-62428-0_32.
- [3] Сушкова, О.С. Применение метода анализа всплескообразной электрической активности мозга для выявления особенностей ЭЭГ пациентов на ранней стадии болезни Паркинсона / О.С. Сушкова, А.А. Морозов, А.В. Габова, А.В. Карабанов // *Журнал неврологии и психиатрии им. С.С. Корсакова*. – 2018. – Т. 118, № 7. – С. 45-48. DOI: 10.17116/jnevro20181187145.
- [4] Sushkova, O.S. Investigation of Specificity of Parkinson's Disease Features Obtained Using the Method of Cerebral Cortex Electrical Activity Analysis Based on Wave Trains / O.S. Sushkova, A.A. Morozov, A.V. Gabova // *13th International Conference on Signal-Image Technology and Internet-Based Systems, India*. – Jaipur: MNIT, 2017. – P. 168-172. DOI: 10.1109/SITIS.2017.37.
- [5] Sushkova, O.S. Investigation of Surface EMG and Acceleration Signals of Limbs' Tremor in Parkinson's Disease Patients Using the Method of Electrical Activity Analysis Based on Wave Trains / O.S. Sushkova, A.A. Morozov, A.V. Gabova, A.V. Karabanov // *Advances in Artificial Intelligence: Proceedings 16th Ibero-American Conference on AI, IBERAMIA, Trujillo, Peru*. – Springer Nature Switzerland AG. – 2018. – Vol. 11238 LNAI. – P. 253-264. DOI: 10.1007/978-3-030-03928-8_21.
- [6] Sushkova, O.S. Investigation of the multiple comparisons problem in the analysis of the wave train electrical activity of muscles in Parkinson's disease patients / O.S. Sushkova, A.A. Morozov, A.V. Gabova, A.V. Karabanov // *Journal of Physics: Conference Series*. – 2019. – Vol. 1368(5). – P. 052004. DOI: 10.1088/1742-6596/1368/5/052004.
- [7] Sushkova, O.S. An Investigation of Accelerometer Signals in the 0.5-4 Hz Range in Parkinson's Disease and Essential Tremor Patients / O.S. Sushkova, A.A. Morozov, A.V. Gabova, A.V. Karabanov, L.A. Chigaleychik // *Advances in Intelligent Systems and Computing: Proceedings of International Conference on Frontiers in Computing and Systems (COMSYS), Jalpaiguri Government Engineering College, West Bengal, India*. – 2021. – Vol. 1255. – P. 455-462. DOI: https://doi.org/10.1007/978-981-15-7834-2_43.

Визуальная среда разработки и визуальное программирование как эффективный инструмент для работы с данными

В.Ф. Зинатуллин
Уфимский государственный нефтяной технический
университет
Уфа, Россия
zinnatullin.vadim2001@yandex.ru

С.Н. Коледин
Уфимский государственный нефтяной технический
университет
Уфа, Россия
koledinsrg@gmail.com

Аннотация—Современный человек сталкивается с большими массивами данных, которые имеют свойство быстрой изменчивости и малопригодны для прогноза. Сложность анализа разноплановой, слабоструктурированной и зачастую не связанной информации требует использования специального программного обеспечения и соответствующих навыков. Предлагается применение визуального программирования для работы с такими данными. В данной работе разработан прототип программного обеспечения для получения, анализа и обработки данных с веб-сайтов в виде визуального языка программирования. Анализируется эффективность и оптимальный функционал визуального программирования для работы с данными.

Ключевые слова— визуальное программирование, язык программирования, работа с данными, веб-скрейпинг, анализ данных.

1. ВВЕДЕНИЕ

В современном мире работа с данными очень востребована. Из открытых веб-источников при использовании специальных программных средств возможно собрать данные, а затем обработать их, что в конечном итоге даёт полезную информацию.

Работа с данными является задачей трудоемкой и требующей навыки программирования выше среднего [1]. Наличие простого и эффективного программного продукта для работы с данными, по сложности не превышающий офисные программы, позволит значительно увеличить круг пользователей при работе с данными. Это, в перспективе, позволит снизить бытовые затраты на удовлетворение базовых потребностей, а также может повысить эффективность малого и среднего бизнеса.

Данная работа ставит перед собой такие задачи:

- исследовать влияния визуального программирования на производительность пользователя в отличие от классического программирования;
- исследовать методы сбора и обработки данных для нахождения оптимального набора функций программного обеспечения, работающего с данными.

2. ВИЗУАЛЬНОЕ ПРОГРАММИРОВАНИЕ. ВЛИЯНИЕ ВИЗУАЛЬНОГО ПРОГРАММИРОВАНИЯ НА ПРОИЗВОДИТЕЛЬНОСТЬ ПОЛЬЗОВАТЕЛЯ

Визуальное программирование – это способ образного, графического представления разрабатываемого алгоритма, который наиболее естественен для восприятия человека [2].

В современном мире визуальное программирование используется в разных сферах деятельности и является востребованным. В отличие от текстового программирования, визуальное программирование может иметь ряд преимуществ, такие как:

- возможность вносить изменения в алгоритм прямо во время его выполнения,
- легкий способ отлаживать алгоритм благодаря визуальному отображению работы программы,
- простота использования.

При исследовании различных сфер применения визуального программирования, таких как: образование (Scratch), робототехника (Lego Mindstorms), разработка видеоигр (Unreal Engine Blueprints, Unity Visual Scripting, Unity Shader Graph), 3D-моделировании (Blender Geometry Nodes), строительство (Dynamo Studio) – выясняется, что в разных сферах визуальное программирование показывает свою эффективность в отличие от текстовых в определенных задачах. Если речь идёт об универсальных языках, то лучше себя проявляют классические языки, однако если рассматривать более узкие сферы, то более эффективным себя может показать визуальное программирование. Так как работа с данными является задачей узконаправленной, то предполагается, что визуальный язык в этой сфере окажется более эффективным средством разработки.

3. МЕТОДЫ СБОРА И ОБРАБОТКИ ДАННЫХ. НАХОЖДЕНИЯ ОПТИМАЛЬНОГО НАБОРА ФУНКЦИЙ ДЛЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Работа с данными включает в себя два компонента: сбор и обработка. Под сбором понимается получение данных с локального диска или с веб-ресурсов. Для сбора данных с веб-ресурсов используется веб-скрейпинг [3].

Обработка данных - процесс превращения данных в информацию. Как правило, цикл обработки данных

состоит из шести основных шагов [4]. Эти шаги описаны далее.

Шаг 1: сбор. Под сбором понимается получение данных с локального диска или с веб-ресурсов.

Шаг 2: подготовка. Подготовка или очистка данных — это процесс сортировки и фильтрации необработанных данных для удаления ненужных и неточных данных. Необработанные данные проверяются на наличие ошибок, дублирования, просчетов или отсутствующих данных и преобразуются в подходящую форму для дальнейшего анализа и обработки. Это сделано для того, чтобы в блок обработки поступали данные только самого высокого качества.

Шаг 3: ввод. На этом этапе необработанные данные преобразуются в машиночитаемую форму и передаются в блок обработки.

Шаг 4: обработка данных. На этом этапе необработанные данные подвергаются различным методам обработки данных, в том числе с использованием алгоритмов машинного обучения и искусственного интеллекта для получения желаемого результата.

Шаг 5: вывод. Данные передаются и отображаются пользователю в удобочитаемой форме, такой как графики, таблицы, векторные файлы, аудио, видео, документы и т. д.

Шаг 6: хранение. Последним этапом цикла обработки данных является хранение, где данные и метаданные хранятся для дальнейшего использования.

Для достижения полного цикла обработки данных визуальная среда программирования должна обеспечивать пользователя необходимыми инструментами для выполнения каждого шага, среди них основные это: инструменты классификации, сортировки, фильтрации, обработки строкового типа данных, а также машинное обучение.

4. РЕАЛИЗАЦИЯ ВИЗУАЛЬНОЙ СРЕДЫ РАЗРАБОТКИ ДЛЯ РАБОТЫ С ДАННЫМИ

Разрабатываемая в работе программа Draft реализовывает в себе два модуля: визуальная среда разработки пользовательского интерфейса и редактор визуального кода. Редактор кода представляет из себя двумерное полотно, в котором можно расположить ноды (блок выполняющий какую-либо функцию) и связывать их друг с другом. Благодаря связям одни ноды передают данные другим нодам, получается своего рода конвейер, который и является алгоритмом конечного приложения. Прототип редактора визуального кода представлен на Рис. 1.

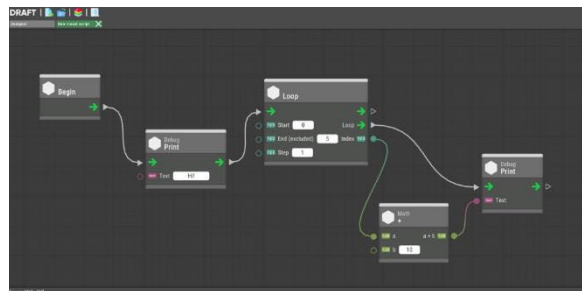


Рис.1. Модуль «Редактор визуального кода» программы Draft

5. ЗАКЛЮЧЕНИЕ

Проведен анализ эффективности различных визуальных языков программирования из разных сфер деятельности. Наибольшую эффективность визуальное программирование показывает в узконаправленных задачах. Так как обработка и сбор данных является узконаправленной задачей, то предполагается, что визуальное программирование в этой сфере будет более эффективным, чем классическое программирование.

Исследован функционал необходимый подобному инструменту для работы с данными, выделен перечень необходимых функций: инструменты классификации, кластеризации, сортировки, фильтрации, обработки строкового типа данных, а также машинное обучение.

Разработан прототип программного обеспечения Draft включающий в себя среду визуальной разработки пользовательских интерфейсов и редактор визуальных скриптов. Целевым функционалом программы является сбор и обработка данных. Конечной задачей является предоставить широкому кругу потребителей простой в освоении и удобное в эксплуатации программное обеспечение для работы с данными.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Фонда содействия развитию малых форм предприятий в научно-технической сфере по программе «УМНИК» (проект № 67696).

ЛИТЕРАТУРА

- [1] Zinnatullin, V.F. Analysis of scientists work directions based on natural language processing and clustering / V.F. Zinnatullin, S.N. Koledin // Proceedings of the VI International conference Information Technology and Nanotechnology. Session Data Science. – 2020. – P. 57-61.
- [2] Коварцев, А.Н. Методы и технологии визуального программирования: учебное пособие / А.Н. Коварцев, В.В. Жидченко, Д.А. Попова-Коварцева. – Самара: Самарский национальный исследовательский университет имени академика С.П. Королёва, 2017. – 200 с.
- [3] Saurkar, A.V. An overview on web scraping techniques and tools / A.V. Saurkar, K.G. Pathare, S.A. Gode // International Journal on Future Revolution in Computer Science & Communication Engineering. – 2021. – Vol. 4(4). – P. 363-367.
- [4] What Is Data Processing: Cycle, Types, Methods, Steps and Examples [Electronic resource]. — Access mode: <https://www.simplilearn.com/what-is-data-processing-article> (20.01.2022).

Моделирование и анализ данных конкурентоспособности территорий при переходе к экономике чистой энергии

И.Н. Хаймович
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самарский университет
государственного управления
«Международный институт рынка»
Самара, Россия
kovalek68@mail.ru

В.М. Рамзаев
Самарский университет
государственного управления
«Международный институт рынка»
Самара, Россия
rvm@imi-samara.ru

В.Г. Чумак
Самарский университет
государственного управления
«Международный институт рынка»
Самара, Россия
imi@imi-samara.ru

Аннотация—Целью данного исследования является разработка моделей и методов принятия управленческих решений на основе прогнозирования экологического фактора. К задачам данного исследования относятся: определение параметров развития экологического фактора конкурентоспособности, разработка модели развития экологического фактора с использованием экспертных оценок, исследование анализа чувствительности данного фактора к эффективности инвестиционных проектов для дальнейшей разработки ПО с применением технологии BIG DATA, были выполнены. К практическим результатам можно отнести повышение качества и своевременности принятия решений по управлению территориями на основе модели управления экологическим фактором модели конкурентоспособности.

Ключевые слова— конкурентоспособность, муниципальные образования, регрессионные модели.

1. ВВЕДЕНИЕ

В настоящее время осуществляется переход мировой экономики к новой экономике чистой энергии. Данный переход осуществляется через инвестирование в ESG – проекты, то есть проекты, связанные с инвестициями в окружающую среду, социальное развитие, корпоративное управление. ESG – проекты приведут к улучшению экологии и к созданию альтернативной энергетике [1]. Для оценки влияния экологического фактора на конкурентоспособность муниципального образования (МО) была создана модель с набором факторов [2]. В данном исследовании проведем моделирование и углубленный анализ данных параметров экологического фактора конкурентоспособности для оценки и прогнозирования развития экологического фактора при увеличении ESG – инвестирования территорий.

Для проведения моделирования зависимости экологических факторов при моделировании конкурентоспособности муниципальных образований применяется частотный анализ анкетных данных по ответам экспертов, но частотный анализ результатов анкеты не позволяет отсеять малозначительные факторы, не устанавливает устойчивые взаимосвязи между ответами анкеты, не раскрывает скрытые зависимости [3-5]. Если анкета сформулирована с большим количеством вопросов (факторов), то частотный анализ размывает и искажает выводы в социологическом исследовании.

Для рассматриваемого исследования потребуется другая методика, которая позволит получить модель, адекватно описывающую экологические факторы для функции конкурентоспособности региона [6-8].

На пути использования методов моделирования в социологическом исследовании встают весьма значительные трудности. Это связано с тем, что социальные явления чрезвычайно сложны, определяются огромным числом различных факторов. Выбор системы нужных факторов, позволяющих объяснить и прогнозировать исследуемые явления, - одна из важнейших проблем в социальном моделировании.

К настоящему времени разработано множество моделей, потенциально применимых в социологическом исследовании. Необходим выбор тех из них, которые наиболее адекватны изучаемым социальным явлениям. Встают проблемы интерпретации результатов моделирования и оценки их надежности.

2. РАЗРАБОТКА МОДЕЛИ ЭКОЛОГИЧЕСКОГО ФАКТОРА КОНКУРЕНТОСПОСОБНОСТИ МУНИЦИПАЛЬНЫХ ОБРАЗОВАНИЙ С АНАЛИЗОМ ЧУВСТВИТЕЛЬНОСТИ ПАРАМЕТРОВ

Для дальнейшего исследования рассматриваем модель экологического фактора конкурентоспособности муниципальных образований Самарской области.

Методика моделирования связей заключается в выявлении методом факторного анализа статистически значимых факторов (корреляционно-связанных) и последующего ранжирования этих факторов по степени влияния на отклик – экологический фактор конкурентоспособности региона.

Базовым фактором является фактор EF, определяющий оценку экологического фактора конкурентоспособности Самарского региона. Предполагаемыми зависимыми переменными (13 параметров) являются: K_p - уровень концентрации промышленного производства в мо; V_p - уровень вредности промышленного производства в мо; T_o - уровень техногенного воздействия объектов мо; A_s - уровень развития автотранспортной сети в мо; Y_n - уровень населенности мо; P_r - обеспеченность мо продуктопроводами (включая нефте- и газопроводы); O_t - состояние очистных сооружений в мо; E_p - уровень

энергоёмкости производства в мо; I_t - использование в мо инновационных технологий переработки отходов; S_b - уровень бюджетного финансирования (за счет собственных средств и поступлений из бюджетов всех уровней) затрат на реализацию природоохранных проектов в мо.

К экологическому фактору можно отнести следующие параметры: уровень развития коммунального хозяйства, степень износа промышленного оборудования, внедрение экологически безопасных видов транспорта и топлива, но эти параметры лучше рассматривать в других факторах конкурентоспособности. Параметр «уровень развития коммунального хозяйств» надо рассматривать в промышленном факторе, степень износа промышленного оборудования и внедрение экологически безопасных видов транспорта и топлива в инфраструктурном факторе.

Факторный анализ был выполнен при помощи статистического пакета SPSS12. В результате были обнаружены связанные параметры, которые можно группировать в факторы, которые имеют разную степень значимости.

Статистическая значимость факторов определялась по коэффициентам корреляции Пирсона и ошибке выборки. В итоге модель экологического фактора конкурентоспособности МО примет вид:

$$EF = \xi_1 * Kp + \xi_2 * Vp + \xi_3 * To + \xi_4 * As + \xi_5 * Yn + \xi_6 * Pr + \xi_7 * Ot + \xi_8 * Ep + \xi_9 * It + \xi_{10} * Sb$$

где ξ - коэффициент значимости группы факторов (определяется экспертным путем посредством анкетирования).

Для дальнейшего исследования рассматриваем модель экологического фактора конкурентоспособности муниципального образования как совокупность трех моделей: для крупных городов (группа 1); для городских округов (группа 2); для муниципальных районов и сельских поселений (группа 3). Методика моделирования связей заключается в выявлении методом факторного анализа статистически значимых параметров (корреляционносвязанных) и последующего ранжирования этих параметров по степени влияния на отклик – экологический фактор конкурентоспособность региона.

3. ЗАКЛЮЧЕНИЕ

Рекомендации по анализу модели экологического фактора конкурентоспособности для МО следующие.

1) В группе экологических факторов важными являются: концентрация промышленного производства; наличие вредных производств; уровень развития автотранспортной сети; уровень развития коммунального хозяйства; наличие на территории нефте-, газо- и продуктопроводов; внедрение современных очистных технологий; внедрение инновационных технологий переработки отходов; внедрение экологически безопасных видов транспорта и топлива; объем средств из бюджета на природоохранные мероприятия (Табл.1).

2) Респонденты исключили из рассмотрения параметр «плотность территории». В динамической модели экологического фактора конкурентоспособности необходимо менять: уровень развития автотранспортной сети, коммунального хозяйства, развитие продуктопроводов, инновационные технологии переработки отходов, объем средств из бюджетов, остальные параметры являются постоянными величинами.

ТАБЛИЦА 1. ЭКОЛОГИЧЕСКИЙ ФАКТОР

Параметры	Важность
Концентрация промышленных предприятий	5
Наличие вредных производств	5
Отсутствие объектов техногенного воздействия	3
Уровень развития автотранспортной сети	5
Уровень развития коммунального хозяйства	5
Плотность населения	3
Наличие на территории нефте-, газо- и продуктопроводов	5
Степень износа промышленного оборудования	4
Внедрение современных очистных технологий и реконструкция действующих очистных сооружений	5
Уровень энергоёмкости производства	4
Внедрение инновационных технологий переработки отходов	5
Внедрение экологически безопасных видов транспорта и топлива	5
Объем средств, выделяемых из бюджетов всех уровней на реализацию природоохранных проектов	5

При дальнейших исследованиях для разработки модели прогнозирования развития экологического фактора конкурентоспособности территорий должны использоваться большие объемы потоковых данных в режиме реального времени. Цель данного исследования, связанная с разработкой моделей и методов принятия управленческих решений на основе прогнозирования экологического фактора была достигнута. Задачи данного исследования, к которым относились: определение параметров развития экологического фактора конкурентоспособности, разработка модели развития экологического фактора с использованием экспертных оценок, исследование анализа чувствительности данного фактора к эффективности инвестиционных проектов для дальнейшей разработки ПО с применением технологий BIG DATA, были выполнены. К практическим результатам можно отнести повышение качества и своевременности принятия решений по управлению территориями на основе модели управления экологическим фактором модели конкурентоспособности.

ЛИТЕРАТУРА

- [1] URL: <https://trends.rbc.ru/trends/green/614b224f9a7947699655a435>.
- [2] Ramzaev, V.M. Modeling and Data Analysis for Assessing the Sensitivity of Competitive Performance of Territories in the Implementation of Investment Projects / V.M. Ramzaev, I.N. Khaimovich, P.V. Chumak // IEEE Xplore. – 2021. DOI: 10.1109/ITNT52450.2021.9649302.

- [3] Ramzaev, V.M. Use of big data technology in public and municipal management / V.M. Ramzaev, I.N. Khaimovich, P.V. Chumak. // CEUR Workshop Proceedings. – 2016. – Vol.1638. – P. 864-872.
- [4] Ramzaev, V.M. Big data analysis for demand segmentation of small business services by activity in region / V.M. Ramzaev, I.N. Khaimovich, P.V. Chumak // CEUR Workshop Proceedings. – 2017. – Vol. 1903. – P. 48-53.
- [5] Ramzaev, V.M. Data modeling for analysis of readiness of municipal education in industry 5.0 / V.M. Ramzaev, I.N. Khaimovich, P.V. Chumak // CEUR Workshop Proceedings. – 2020. – Vol. 2667. – P. 1-4.
- [6] Geraskin, M. Model and algorithm of industrial risk control at regional level / M. Geraskin, E. Rostova // CEUR Workshop Proceedings. – 2020. – Vol. 2667. – P. 1-4.
- [7] Geraskin, M. Analysis of monopolistic competition in consumer goods markets with credit sales / M. Geraskin, O. Kuznetsova // CEUR Workshop Proceedings. – 2020. – Vol. 2667. – P. 1-4.
- [8] Geras'kin, M.I. Analysis of Game-Theoretic Models of an Oligopoly Market under Constrains on the Capacity and Competitiveness of Agents / M.I. Geras'kin, A.G. Chkhartishvili // Automation and Remote Control. – 2017. – Vol. 78(11). – P. 2025-2038.
- [9] Komarevtseva, O.O. Simulation of data for determining the readiness of municipalities to implement smart city technologies / O.O. Komarevtseva // DAMDID/RCDL. – 2017. – P. 167-177.

Пространственно-временной скейлинг зрительно-вызванных нейромагнитных сигналов человека

В.А. Юнусов
Казанский (Приволжский)
федеральный университет
Казань, Россия
valentin.yunusov@gmail.com

С.А. Демин
Казанский (Приволжский)
федеральный университет
Казань, Россия
serge_demin@mail.ru

И.А. Русанова
Казанский (Приволжский)
федеральный университет
Казань, Россия
irusanova@yandex.ru

А.В. Минкин
Казанский (Приволжский)
федеральный университет
Казань, Россия
minkins@yandex.ru

А.А. Еленёв
Казанский (Приволжский)
федеральный университет
Казань, Россия
a.elenev6345@gmail.com

Аннотация—В настоящей работе проводится фрактальный анализ зрительно-вызванных нейромагнитных откликов коры головного мозга человека в ответ на мерцающие световые стимулы разных цветовых комбинаций (красно-голубой, красно-зеленый, голубой-зеленой). В исследовании используются сигналы магнитоэнцефалограмм здоровых испытуемых, фиксируемые до и после воздействия стимулов. Для расчета фрактальной размерности D используется модифицированный метод Хигучи. Пространственно-временной скейлинг демонстрируется на основе топографических карт и гистограмм. Показано, что включение световых стимулов приводит к значительной реакции со стороны практически всех отделов коры головного человека. Наиболее значительное изменение самоподобных свойств исследуемых сигналов наблюдается для лобной и затылочной областей. Полученные результаты будут представлять интерес для наук о данных, нейрофизиологии и когнитивной психологии.

Ключевые слова— науки о данных, анализ временных сигналов, фрактальные особенности, нейромагнитные отклики, мерцающие световые стимулы, когнитивные возможности, диагностика.

1. ВВЕДЕНИЕ

Начиная с 40-х годов 20 века в человеческой науке происходит эволюция наук о сложности согласно пяти интеллектуальным традициям: теории динамических систем, общей теории систем, кибернетике, искусственному интеллекту и теории сложных систем. Благодаря взаимному проникновению этих направлений и широкому распространению полученных результатов в научном сообществе становится возможным изучение уникальных свойств сложных как физических, так и нефизических систем. Настоящим триумфом наук о сложности стало присуждение в 2021 году Нобелевской премии в области физики троим ученым: Сюкуро Манабэ, Клаусу Хассельману и Джордžo Паризи «...за фундаментальный вклад в наше понимание сложных физических систем». В работах Манабэ приводится математическое описание сложных изменчивых систем на примере исследования климата Земли [1], в то время как Хассельману удалось использовать стохастические методы для обнаружения связи между постоянно

меняющимися погодными условиями и более устойчивым климатом Земли [2]. Паризи удалось применить статистические методы и модели для установления взаимосвязей в хаосе и флуктуациях в сложных физических системах от атомарных до планетарных масштабов [3].

Анализируя научную литературу, можно выделить такие свойства сложных систем как сложность, открытость, нелинейность, адаптивность, эмерджентность, самоорганизацию и критические переходы [4]. Под сложной системой понимается объект, состоящий из значительного числа взаимодействующих элементов. Взаимодействие между частями целого приводит к перечисленным уникальным свойствам, которые делают системы «сложными» для изучения классическими методами. Значительный прогресс в области вычислительной техники и экспериментального регистрирующего оборудования способствовал привлечению к изучению свойств сложных систем наук о данных. Анализ временных сигналов численными методами позволяет с большей точностью изучить особенности динамических взаимосвязей, которые выстраиваются в ходе эволюции сложных систем.

2. ФРАКТАЛЬНЫЕ ОСОБЕННОСТИ НЕЙРОМАГНИТНЫХ ОТКЛИКОВ ЧЕЛОВЕКА

Ранее с целью диагностики фоточувствительной эпилепсии (ФЧЭ) в работах авторов [5, 6] были представлены результаты исследования магнитоэнцефалограмм (МЭГ) коры головного человека в ответ на мерцающие световые стимулы разных цветовых комбинаций. В дополнение к установленным диагностическим критериям в работе [7] изучены фрактальные закономерности сигналов МЭГ при ФЧЭ.

В настоящей работе мы развиваем предложенный ранее подход к расчету фрактальной размерности D для анализа МЭГ сигналов здоровых испытуемых в ответ на красно-голубой, красно-зеленый и голубой-зеленый стимулы, основанный на модифицированном методе Хигучи. Экспериментальные данные были зафиксированы 61 СКВИДами (сверхпроводящий квантовый интерференционный датчик), расположенными на всей поверхности головы человека.

Примерные схемы расположения датчиков представлены на Рис. 1, Рис. 2.

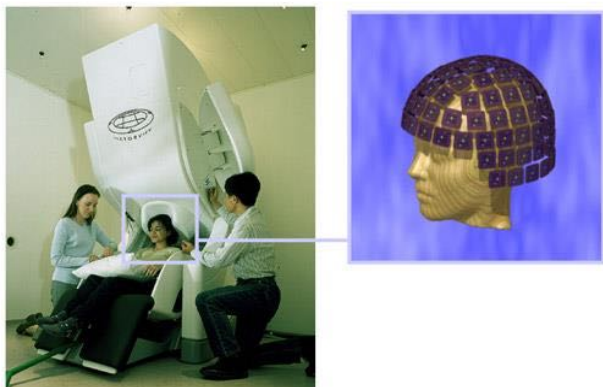


Рис. 1. Регистрация нейромагнитных откликов человека и примерное расположение СКВИДов на поверхности головы

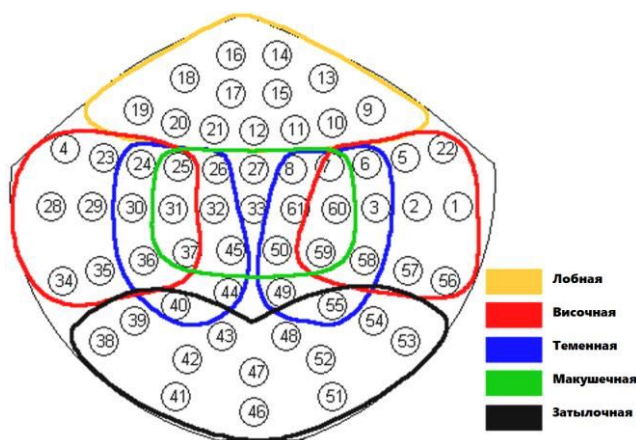


Рис.2. Локализация СКВИДов на поверхности головы человека

Регистрация сигналов выполнялась для семи мужчин и двух женщин в возрасте от 22 до 27 лет [8]. Ни у кого из испытуемых не были установлены какие-то неврологические или другие отклонения в работе головного мозга. Световые стимулы подавались на экран при помощи двух проекторов. Первые 200 измерений фиксировался контрольный сигнал: испытуемые смотрели на экран без подачи стимула, последующие 201–1095 измерений выполнялись при подаче стимула.

Результаты исследования представлены в виде топографических карт, где приведены значения показателя D для каждой локализации СКВИДа, а также в виде гистограмм, для которых рассчитаны усредненные значения D для отдельных областей коры головного мозга человека.

Результаты настоящего исследования позволили установить, каким образом меняется фрактальная структура МЭГ сигналов в ответ на разные световые стимулы. В частности, подача мерцающего стимула, то есть переход от спонтанных к индуцированным или вызванным нейромагнитным откликам, приводит к изменению пространственно-временного скейлинга сигналов в большинстве областей коры головного мозга здоровых испытуемых. Наибольшая реакция выявлена для лобной и затылочной областей.

3. ЗАКЛЮЧЕНИЕ

Изучение эффектов пространственно-временной инвариантности осуществляется на основе фрактальных методов, которые позволяют определить различные значения и распределения фрактальной и мультифрактальной размерностей, степенные законы с целочисленными и дробными показателями, меры сложности временных сигналов [9].

В настоящей работе проводится фрактальный анализ сигналов МЭГ человека при воздействии мерцающими световыми стимулами разных цветовых комбинаций. В ходе исследования установлены области коры головного мозга с наибольшей реакцией на внешнее воздействие. Работа будет интересна с точки зрения биофизики восприятия визуальной информации и развития когнитивных возможностей человека. Кроме того, результаты настоящей работы будут полезны для изучения механизмов провоцирования приступов при фоточувствительной эпилепсии [5–8, 10].

Дальнейшие исследования самоподобной структуры МЭГ сигналов будут проводиться путем их комплексного анализа с привлечением других фрактальных показателей.

БЛАГОДАРНОСТИ

Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»).

ЛИТЕРАТУРА

- [1] Manabe, S. The Effects of Doubling the CO₂ Concentration on the climate of a General Circulation Model / S. Manabe, R.T. Wetherald // *Journal of the Atmospheric Sciences*. – 1975. – Vol. 32. – P. 3-15.
- [2] Hasselmann, K. Multi-pattern fingerprint method for detection and attribution of climate change / K. Hasselmann // *Climate Dynamics*. – 1997. – Vol. 13. – P. 601-611.
- [3] Altarelli, G. Asymptotic freedom in parton language / G. Altarelli, G. Parisi // *Nuclear Physics B*. – 1977. – Vol. 126(2). – P. 298-318.
- [4] Weaver, W. Science and Complexity / W. Weaver // *American Scientist*. – 1948. – Vol. 36(4). – P. 536-44.
- [5] Panishev, O.Y. Cross-correlation markers in stochastic dynamics of complex systems / O.Y. Panishev, S.A. Demin, J. Bhattacharya // *Physica A*. – 2010. – Vol. 389. – P. 4958-4969.
- [6] Timashev, S.F. Frequency and Phase Synchronization in Neuromagnetic Cortical Responses to Flickering-Color Stimuli / S.F. Timashev, Y.S. Polyakov, R.M. Yulmetyev, S.A. Demin, O.Y. Panishev, S. Shimojo, J. Bhattacharya // *Laser Physics*. – 2010. – Vol. 20(3). – P. 604-617.
- [7] Panishev, O.Y. Studying Properties of Abnormal Human Brain Activity in Photosensitive Epilepsy Caused by Light Stimulation / O. Y. Panishev, S.A. Demin, I.A. Rusanova // *Biomedical Engineering*. – 2015. – Vol. 49(4). – P. 236-239.
- [8] Bhattacharya, J. Nonlinear dynamics of evoked neuromagnetic responses signifies potential defensive mechanisms against photosensitivity / J. Bhattacharya, K. Watanabe, S. Shimojo // *International Journal of Bifurcation and Chaos*. – 2004. – Vol. 14(8). – P. 2701-2720.
- [9] Bhagat, M. Investigating Neuromagnetic Brain Responses against Chromatic Flickering Stimuli by Wavelet Entropies / M. Bhagat, C. Bhushan, G. Saha, S. Shimjo, K. Watanabe, J. Bhattacharya // *PLoS One*. – 2009. – Vol. 4. – P. e7173.
- [10] Poleon, S. Photosensitivity in generalized epilepsies / S. Poleon, J.P. Szaflarski // *Epilepsy & Behavior*. – 2017. – Vol. 68. – P. 225-233.

Мультиклассовая классификация для диагностики функционирования сложных технических систем

Ю.Е. Кувайскова
Ульяновский государственный
технический университет
Ульяновск, Россия
u.kuvaiskova@mail.ru

Н.А. Ломовцева
Ульяновский государственный
технический университет
Ульяновск, Россия
natalya.lomovtseva@gmail.com

В.Н. Клячкин
Ульяновский государственный
технический университет
Ульяновск, Россия
v_kl@mail.ru

Abstract—Диагностика функционирования сложных технических систем часто сводится к разделению состояния объекта на несколько классов: исправное состояние и различные типы неисправностей. С помощью разработанной программы проводится мультиклассовая классификация методами машинного обучения. В качестве примера рассмотрены особенности диагностики компьютерной системы.

Keywords—*диагностика, классификация, кросс-валидация, тестовая выборка, F-мера.*

1. ВВЕДЕНИЕ

Диагностика функционирования сложных технических систем проводится в целях обеспечения их безопасности и надежности [1]. При этом часто задача сводится к разделению состояния объекта на несколько классов: исправное состояние и различные типы неисправностей. В качестве исходных данных используются результаты наблюдений за функционированием объекта в период предшествующей эксплуатации: набор показателей работы объекта и его состояние.

Задача сводится к мультиклассовой классификации методами машинного обучения [2-3]. По имеющемуся набору исходных данных строится классификатор, с помощью которого можно спрогнозировать состояние объекта при новом наборе показателей.

Частный случай такой задачи при классификации режимов работы авиационного газотурбинного двигателя решалась с применением нейронных сетей [4]. Еще один частный случай – бинарная классификация, когда состояние объекта разделяется на два класса – исправное и неисправное, был использован для вибродиагностики гидроагрегата [5].

Необходимо выбрать метод классификации, обеспечивающий нужную точность, провести отбор значимых показателей, задать объем контрольной выборки. Для исключения переобучения может быть использована кросс-валидация. Это вопросы, решаемые в любой задаче машинного обучения.

Особенностью диагностики сложных технических систем является сравнительно небольшой объем наблюдений и несбалансированность классов: как правило наблюдений при неисправных состояниях объекта гораздо меньше, чем при исправных. В этом случае основной критерий качества обучения – доля

правильных ответов на тестовой выборке – может привести к ложным выводам.

2. ПРОГРАММА МУЛЬТИКЛАССОВОЙ КЛАССИФИКАЦИИ

Программа разработана на языке программирования Python. Обеспечивается ввод файла исходных данных, разбиение его в заданном соотношении на обучающую и тестовую части случайным образом с применением стратификации, проведение кросс-валидации с заданным пользователем количеством блоков. Программа обеспечивает также отбор значимых показателей функционирования объекта двумя методами: по коррелированности показателей и по значимости регрессоров в линейной регрессионной модели зависимости номера класса объекта от показателей.

Используется библиотека Sklearn, из которой импортируются готовые конструкторы для базовых классификаторов LogisticRegression, LinearDiscriminantAnalysis, KNeighborsClassifier, DecisionTreeClassifier, GaussianNB, SVC (метод опорных векторов), а также композиционных классификаторов RandomForestClassifier (случайный лес) и AdaBoostClassifier (адаптивный бустинг).

Для реализации агрегирования по среднему значению использованы функции Python для нахождения вероятностей принадлежности каждого объекта к соответствующему классу: `predict_proba`.

Из этой же библиотеки Sklearn импортируются модули для вычисления метрик мультиклассовой классификации: `accuracy_score` – функционал ошибок; `f1_score` – F-мера (основная характеристика качества классификации для несбалансированных классов) и `confusion_matrix` – матрица неточностей, а также модуль для кросс-валидации: `cross_val_score` и модуль для разбиения исходных данных на обучающую и тестовую выборки: `train_test_split`.

Результат работы программы – расчет трех мер качества классификации по тестовой выборке для каждого из базовых, композиционных и агрегированных классификаторов.

Пользователь в зависимости от характера исходных данных выбирает наилучший классификатор по тому или иному критерию. Этот классификатор в дальнейшем используется для прогнозирования состояния объекта. По вновь найденным показателям функционирования объекта прогнозируется, к какому классу относится его состояние.

3. ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ

Исследовалось влияние различных показателей работы компьютера (загрузки процессора, температуры нагрева компонентов динамической памяти, напряжения, мощности и других – всего 9 показателей $x_1 \dots x_9$) на его состояние: 1 – исправен, 2 – зависание, 3 – искажение изображения, 4 – заторможенность курсора, 5 – заедание звука. Всего получено 715 наблюдений, из которых нарушения имели место в 143 случаях (20%).

Исследовалось влияние отбора показателей, доли тестовой выборки и метода обучения на качество классификации по F-мере на тестовой части выборки. В рассматриваемом примере отбор значимых показателей несущественно влияет на результат, максимальное значение рассматриваемого критерия получено при объеме тестовой выборки 25% и методе классификации – дерево решений – таблица 1).

Увеличение объема тестовой выборки до 30% не улучшило результат: в этом случае высшее значение F-меры на тестовой выборке составило 0,973 при агрегировании случайного леса с адаптивным бустингом.

Заметим, что в отличие от бинарной классификации, где агрегированные методы практически всегда улучшали результаты, в мультиклассовой задаче это обстоятельство не подтвердилось.

Варьирование объема тестовой выборки не увеличило заметно качество прогнозирования состояния компьютера (расхождение составляет 2-3%). Такого порядка расхождение имеет место при повторении испытаний (за счет случайного отбора данных в обучающую и тестовую выборки).

Таблица 1. F-МЕРА ПРИ РАЗЛИЧНЫХ БАЗОВЫХ И КОМПОЗИЦИОННЫХ КЛАССИФИКАТОРАХ С ОБЪЕМОМ ТЕСТОВОЙ ВЫБОРКИ 25%

Метод классификации	F-мера
Дерево решений	0,978
Случайный лес	0,960
Метод опорных векторов	0,906
Метод K-ближайших соседей	0,909
AdaBoost	0,823
Логистическая регрессия	0,923
Линейный дискриминантный анализ	0,899

Следует отметить, что при разных объемах тестовой выборки наилучший результат показывают разные методы обучения. При этом в двух из пяти испытаний применение агрегированных классификаторов [6-8] повысило качество диагностики.

4. ЗАКЛЮЧЕНИЕ

Разработанная программа мультиклассовой классификации для диагностики функционирования сложных технических систем методами машинного

обучения обеспечивает достаточно высокую точность классификации по предварительно сформированной выборке исходных данных по результатам предшествующей работы объекта.

Наилучший классификатор может быть использован для прогноза состояния системы по заданным показателям ее функционирования.

В рассмотренном примере диагностики функционирования компьютерной системы выявлено, что отбор значимых факторов и изменение доли тестовой выборки не приводят к значимому повышению качества классификации, однако в других сложных технических системах это влияние может быть существенным, поэтому разработанная программа предусматривает возможность учета и варьирования этих факторов.

В общем случае алгоритм диагностики функционирования сложных технических систем путем мультиклассовой классификации должен включать:

- исследование влияния отбора значимых показателей,
- оценка наилучшей доли тестовой выборки,
- оценка количества блоков кросс-валидации,
- влияние метода машинного обучения на значение выбранного критерия качества диагностики.

При этом основным критерием качества диагностики служит доля правильных ответов на тестовой выборке (при достаточно большом количестве данных о неисправном состоянии объекта) или F-мера при несбалансированных данных.

ЛИТЕРАТУРА

- [1] Биргер, И.А. Техническая диагностика / И.А. Биргер. – М.: Машиностроение, 1978. – 240 с.
- [2] Wang, B. A Multicategory Kernel Distance Weighted Discrimination Method for Multiclass Classification / B. Wang, H. Zou // *Technometrics*. – 2019. – Vol. 61(3). – P. 396-408.
- [3] Paynabar, K. A Change-Point Approach for Phase-I Analysis in Multivariate Profile Monitoring and Diagnosis / K. Paynabar, Ch. Zou, P. Qiu // *Technometrics*. – 2016. – Vol. 58(2). – P. 191-204.
- [4] Васильев, В.И. Классификация режимов работы ГТД с использованием технологии нейронных сетей / В.И. Васильев, С.В. Жернаков // *Вестник Уфимского государственного авиационного технического университета*. – 2009. – Т. 12, № 1. – С. 53-56.
- [5] Клячкин, В.Н. Прогнозирование и диагностика стабильности функционирования технических объектов: монография / В.Н. Клячкин, В.Р. Крашенинников, Ю.Е. Кувайскова. – М.: Русайнс, 2020. – 200 с.
- [6] Klyachkin, V.N. Aggregated classifiers for state diagnostics of the technical object / V.N. Klyachkin, J.E. Kuvaykova, D.A. Zhukov // *International Multi-Conference on Industrial Engineering and Modern Technologies*. – 2019. – P. 8934362.
- [7] Klyachkin, V.N. Aggregated Classifiers for State Diagnostics of the Technical Objects International Multi-Conference on Industrial Engineering and Modern Technologies / V.N. Klyachkin, J.E. Kuvaykova, D.A. Zhukov // *FarEastCon*. – 2019. – Vol. 2019. – P. 8934362.
- [8] Zhukov, D.A. Selection of Aggregated Classifiers for the Prediction of the State of Technical Objects / D.A. Zhukov, V.N. Klyachkin, V.R. Krashennnikov, Yu.E. Kuvaykova // *CEUR Workshop Proceedings*. – 2019. – P. 361-365.

Использование машинного обучения для управления навигацией робота по показаниям датчиков расстояний

Н.А. Ломовцева
Ульяновский государственный
технический университет
Ульяновск, Россия
natalya.lomovtseva@gmail.com

Ю.Е. Кувайскова
Ульяновский государственный
технический университет
Ульяновск, Россия
u.kuvaiskova@mail.ru

И.Н. Карпунина
Ульяновский институт
гражданской авиации
Ульяновск, Россия
karpunina53@yandex.ru

В.Н. Клячкин
Ульяновский государственный
технический университет
Ульяновск, Россия
v_kl@mail.ru

Д.М. Ястребов
Научно-производственное
объединение «Марс»
Ульяновск, Россия
mars@mv.ru

Abstract—По показаниям ультразвуковых датчиков для измерения расстояний, расположенных на роботе, требуется спрогнозировать направление перемещения робота. Это задача мультиклассовой классификации, решаемая методами машинного обучения. Решение проводилось в системе Statistica с применением метода Random Forest.

Keywords— мультиклассовая классификация, перемещение робота, система Statistica, Random forest.

1. ВВЕДЕНИЕ

По показаниям ультразвуковых датчиков для измерения расстояний, расположенных на роботе, требуется спрогнозировать направление перемещения робота. Предполагается наличие неизвестной зависимости между показаниями датчиков и перемещениями робота. По результатам предварительных испытаний известна совокупность пар: набору показателей X соответствует состояние Y – это исходная выборка результатов наблюдений. По этим данным надо построить алгоритм для восстановления зависимости, с помощью которой по новым данным о показаниях датчиков можно выдать достаточно точный ответ о направлении его перемещения. Это задача мультиклассовой классификации, решаемая методами машинного обучения, или обучения по прецедентам (с учителем) [1-3].

Существует множество методов решения этой задачи. Это и базовые статистические методы (наивный байесовский классификатор, дискриминантный анализ, метод опорных векторов и др.), и нейронные сети, и композиционные методы (бэггинг и бустинг). Для бинарной классификации более эффективны агрегированные методы. В настоящей работе решение проводилось в системе Statistica [4] с применением метода Random Forest (случайный лес), показавшего лучшие результаты именно для мультиклассовой классификации.

2. ПРОВЕДЕНИЕ МАШИННОГО ОБУЧЕНИЯ

Случайный лес – один из наиболее эффективных методов машинного обучения, используемых при решении задач классификации и регрессии [2-3]. Это модель, состоящая из множества независимых деревьев решений и использующая две ключевые концепции при построении каждого дерева, которые и делают этот лес случайным:

1. Случайная выборка наблюдений из обучающего набора данных.
2. Случайный набор предикторов из всех известных предикторов при принятии решений о разбиении узлов.

Модуль случайного леса включен в систему Statistica, позволяет варьировать различные факторы, влияющие на качество обучения, и может быть использован для решения поставленной задачи прогнозирования перемещения робота по результатам показаний датчиков расстояний, расположенных на роботе.

Точность прогнозирования случайного леса зависит от ряда факторов – количества предикторов в случайном наборе, объема подвыборки (случайной выборки из обучающего набора), количества деревьев, максимальной глубины деревьев (максимального количества уровней), максимального количества узлов в деревьях, минимального числа объектов в листьях, минимального количества объектов в дочернем узле [6-7].

Для численного исследования использовались данные, полученные на кафедре инженерной телеинформатики Федерального университета штата Сеара в Форталезе (Бразилия) [5]: 24 ультразвуковых датчика для измерения расстояния были расположены в различных частях робота. Фиксировалось направление перемещения робота: отклик Y , характеризующий перемещение робота, представлен номером соответствующего класса: 1 – движение вперед, 2 – плавный поворот направо, 3 – резкий поворот направо, 4 – плавный поворот налево. Всего получено 5456 наблюдений.

Обозначения показаний датчиков: X_1 – датчик в передней части робота, опорный угол -180° ; X_2 – опорный угол -165° ; X_3 – опорный угол -150° ; X_4 – опорный угол -135° ; ... X_{12} – опорный угол -15° ; X_{13} – датчик в задней части робота, опорный угол 0° ; X_{14} – опорный угол 15° ; ... X_{23} – опорный угол 150° ; X_{24} – опорный угол 165° .

Исследовалось влияние количества случайных предикторов, используемых при принятии решений о разбиении узлов деревьев, максимальной глубины деревьев, а также количества деревьев на точность прогнозирования случайным лесом направлений перемещения робота.

Выбор этих факторов в каждой конкретной задаче может существенно улучшить качество классификации.

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

При значениях этих факторов, установленных в системе Statistica по умолчанию, построенная матрица неточностей (рис. 1) показывает, что случайный лес присвоил метку 1-го класса 237 наблюдениям на тестовой выборке, из которых верно было распознано 235, что составляет 99,16%, такой же процент правильно распознанных наблюдений от общего числа наблюдений 1-го класса (движение вперед) на тестовой выборке и 41,89% от всех наблюдений на тестовой выборке.

Случайный лес обнаружил 2-й класс (плавный поворот направо) в 77 случаях, все из которых верные (100%) и составляют 98,72% от общего количества наблюдений 2-го класса на тестовой выборке.

Все наблюдения 3-го класса (резкий поворот направо) на тестовой выборке были распознаны верно, что составляет 100%.

Процент правильно распознанных наблюдений 4-го класса (плавный поворот налево) от общего количества класса на тестовой выборке – 92,86% (рис. 1).

Classification matrix (data.sta)						
Response: Y						
Test set sample; Number of trees: 100						
	Observed	Class Predicted 1	Class Predicted 2	Class Predicted 3	Class Predicted 4	Row Total
Number	1	235		2		237
Column Percentage		99.16%	0.00%	0.90%	0.00%	
Row Percentage		99.16%	0.00%	0.84%	0.00%	
Total Percentage		41.89%	0.00%	0.36%	0.00%	42.25%
Number	2		77	1		78
Column Percentage		0.00%	100.00%	0.45%	0.00%	
Row Percentage		0.00%	98.72%	1.28%	0.00%	
Total Percentage		0.00%	13.73%	0.18%	0.00%	13.90%
Number	3			218		218
Column Percentage		0.00%	0.00%	98.64%	0.00%	
Row Percentage		0.00%	0.00%	100.00%	0.00%	
Total Percentage		0.00%	0.00%	38.86%	0.00%	38.86%
Number	4	2			26	28
Column Percentage		0.84%	0.00%	0.00%	100.00%	
Row Percentage		7.14%	0.00%	0.00%	92.86%	
Total Percentage		0.36%	0.00%	0.00%	4.63%	4.99%
Count	All Groups	237	77	221	26	561
Total Percent		42.25%	13.73%	39.39%	4.63%	

Рис. 1. Матрица неточностей случайного леса (количество случайных предикторов – 5, максимальная глубина – 10, количество деревьев – 100)

На рис. 2 представлен процесс построения случайного леса: по горизонтальной оси отложено количество деревьев, по вертикальной – доля ошибок классификации на обучающей (красная линия) и тестовой (синяя линия) выборках. Видно, что процесс практически стабилизировался при использовании 50 деревьев.

Наилучший результат прогнозирования случайным лесом направлений перемещения робота обеспечивается при построении 50 деревьев с максимальной глубиной – 10, которые принимают решения о разбиении узлов с использованием семи случайных предикторов.

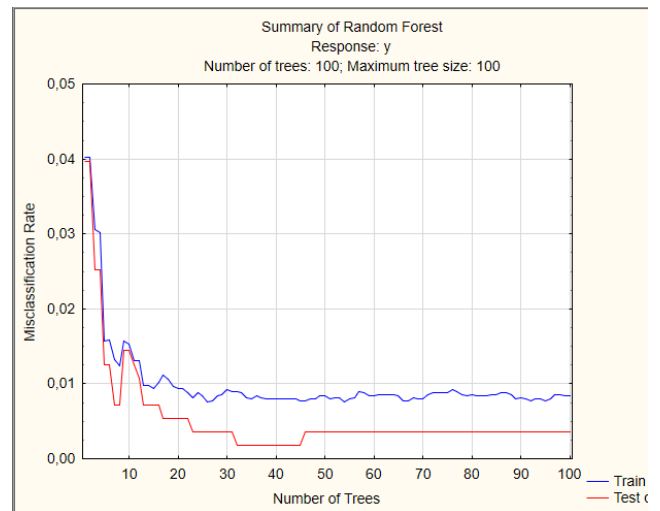


Рис. 2. Построение случайного леса

4. ЗАКЛЮЧЕНИЕ

Проведенное исследование показало, что мультиклассовая классификация в системе Statistica с применением случайного леса обеспечивает достаточно качественное прогнозирование направления перемещений робота по результатам показаний датчиков расстояний.

Рассмотренный подход может быть использован и при диагностике сложных технических систем, когда необходимо оценить по результатам наблюдений за показателями функционирования, в каком из множества состояний находится исследуемый объект.

ЛИТЕРАТУРА

- [1] Witten, I.H. Data mining practical machine learning tools and techniques / I.H. Witten, E. Frank. – SF: Morgan Kaufmann Publ, 2005.
- [2] Клячкин, В.Н. Прогнозирование и диагностика стабильности функционирования технических объектов: монография / В.Н. Клячкин, В.Р. Крашенинников, Ю.Е. Кувайскова // Москва: Русайнс, 2020.
- [3] Wang, B. A Multicategory Kernel Distance Weighted Discrimination Method for Multiclass Classification / B. Wang, H. Zou // Technometrics. – 2019. – Vol. 61(3). – P. 396-408. DOI: 10.1080/00401706.2018.1529629.
- [4] Боровиков, В.П. STATISTICA. Искусство анализа данных на компьютере / В.П. Боровиков. – СПб: Питер, 2003.
- [5] Freire, A.L. Short-Term Memory Mechanisms in Neural Network Learning of Robot Navigation Tasks: A Case Study / A.L. Freire, G.A. Barreto, M. Veloso, A.T. Varela // Proceedings of the 6th Latin American Robotics Symposium (LARS). – Valparaíso-Chile, 2009. – P. 1-6.
- [6] Wyner, A.J. Explaining the success of adaboost and random forests as interpolating classifiers / A.J. Wyner // The Journal of Machine Learning Research. – 2017. – Vol. 18(1). – P. 1558-1590.
- [7] Kuvayskova, Yu. Recognition and Forecasting of a Technical Object State based on its Operation Indicators Monitoring Result / Yu. Kuvayskova, V. Klyachkin, V. Kracheninnikov // International Multi-Conference on Industrial Engineering and Modern Technologies. – 2020. – P. 9271119. DOI: 10.1109/FarEastCon50210.2020.9271119.

Алгоритм выбора тарифов страхования промышленных рисков на основе анализа больших массивов данных об объектах страхования

звание статьи

М.И. Гераскин

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
innovation@ssau.ru

Е.П. Ростова

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
el_rostova@mail.ru

Аннотация—Предлагается решение задачи обработки больших массивов информации об объектах страхования промышленных рисков с целью оптимального формирования системы тарифов страховых компаний, аккумулирующих в портфелях десятки и сотни миллионов застрахованных объектов, дифференцированных по большому числу параметров. Многомерные функции страховых тарифов от параметров застрахованных объектов моделируются на основе оптимизации с учетом вероятности ущерба, зависящей от превентивных мероприятий страхователей.

Ключевые слова— промышленный риск, управление рисками, аварийность, страховой портфель.

1. ВВЕДЕНИЕ

Управление промышленными рисками возможно с помощью такого инструмента, как страхование. В РФ действует добровольное страхование гражданской ответственности владельцев опасных производственных объектов. Данный вид страхования охватывает производственные объекты, эксплуатирующие резервуары под давлением, подъемные механизмы (включая лифты, эскалаторы и траволаторы), гидротехнические сооружения и др. Объем страхового портфеля отдельных компаний превышает десятки миллионов договоров (см. таблица 1), для формирования тарифной ставки обрабатывается информация о десятках параметров по каждому объекту [1].

Таблица 1. Количество застрахованных объектов в 2020 году

Страховая компания	Количество действовавших договоров страхования, млн	Количество заключенных (новых) договоров страхования, млн
ПАО СК «Росгосстрах»	12, 087 299	16, 237 078
САО «РЕСО-Гарантия»	10, 906 426	12, 424 164
АО «Альфастрахование»	16, 949 637	34, 688 211

Проблемы формирования страхового портфеля в отечественных [2], [3] и зарубежных [4], [5] исследованиях рассматриваются с точки зрения финансовых механизмов и методов эконометрического моделирования. В последнем случае исследователи рассматривают различные варианты распределения случайных величин ущерба и частоты наступления

страхового случая. В данном исследовании разработана модель формирования портфеля страхования промышленных рисков с учетом параметров всех застрахованных объектов.

2. МЕТОДЫ

Рассмотрим страховую компанию, имеющую n застрахованных объектов. Плотность вероятности наступления страхового события $p=e^{-f\xi}$ зависит от превентивных мероприятий, проводимых страхователем на сумму f , эффективность которых ξ определяется по результатам статистики об объекте страхования за несколько временных периодов по ряду параметров. Ожидаемый ущерб $X=p\chi$ зависит от вероятности p и от величины ущерба χ . Страховая премия V зависит от доли застрахованного риска γ^S ($X^S = \gamma^S X$) и тарифной ставки T : $V=V(X^S, T)$. Размер страхового возмещения W зависит от ущерба $X=p\chi$ и от уровня ответственности страховщика α : $W=W(X, \alpha)$. Тогда функция прибыли страховщика:

$$\Pi = \sum_{i=1}^n (V_i - p_i W_i) = \quad (1)$$

$$= \sum_{i=1}^n \left(\gamma_i^S e^{-\xi_i f_i(T_i)} \chi_i T_i - e^{-\xi_i f_i(T_i)} \chi_i \alpha_i \right)$$

Тарифная ставка является инструментом страховой компании, с помощью которого, возможно формирование страхового портфеля. В предыдущих работах было показано взаимное влияние тарифной ставки и предупредительных мероприятий по снижению застрахованного риска: чем выше скидка страхователю, тем выгоднее предприятию проводить предупредительные мероприятия. Задача определения вектора оптимальных тарифных ставок T_i^* , $i = \overline{1, n}$ по критерию максимизации прибыли будет иметь вид:

$$\{T_i^*\} = \arg \max_{0 < T_i < 1} \Pi, \quad \begin{cases} \frac{df_i}{dT_i} < 0 \\ 0 < \xi_i < 1 \\ \sum_{i=1}^n (\gamma_i^S T_i - \alpha_i) > 0 \end{cases} \quad (2)$$

Решив поставленную задачу, получим следующий вид функции страхового тарифа:

$$T_i^* = \frac{\xi_i k_i}{2} \left(\sqrt{1 + \frac{4\alpha_i}{\xi_i k_i \gamma_i^S}} - 1 \right) \quad (3)$$

при условии

$$\frac{\xi_i \gamma_i^S k_i + 2\alpha_i}{2} \left[\sqrt{1 + \frac{4\alpha_i}{\xi_i \gamma_i^S k_i}} - 1 \right] - \alpha_i < 0 \quad (4)$$

Поставленная задача (1), (2) позволяет определить страховой тариф с учетом предупредительных мероприятий и параметров всех застрахованных рисков.

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

Предложенная задача (1), (2) была решена для предприятий металлургической и электротехнической отраслей. В таблице 2 представлены функции тарифных ставок для каждого из исследуемых объектов.

ТАБЛИЦА 2. ФУНКЦИЯ T^* ДЛЯ МЕТАЛЛУРГИЧЕСКОГО И ЭЛЕКТРОТЕХНИЧЕСКОГО ПРЕДПРИЯТИЯ

Предприятие	T^*
Металлургическое	$T_M^* = 39 \cdot 10^{-5} \left(\sqrt{1 + 5065,2 \frac{\alpha_M}{\gamma_M^S}} - 1 \right)$
Электротехническое	$T_E^* = 2,9 \cdot 10^{-5} \left(\sqrt{1 + 68610,6 \frac{\alpha_E}{\gamma_E^S}} - 1 \right)$

Построим графики, являющиеся линиями уровня для T_M^* при различных значениях уровня ответственности страховщика α (рисунок 1).

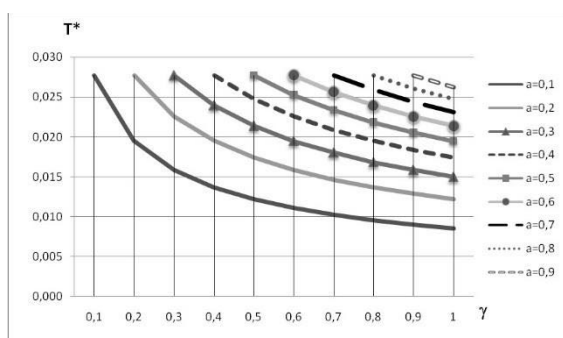


Рис. 1. Зависимость T_M^* от значений α и γ^S

Из графика видно, что максимальный уровень тарифной ставки достигается при $\alpha=0,9$, что соответствует наибольшей доле ответственности страховщика. Также заметно влияние параметра γ^S , отражающего долю застрахованного риска, состоящее в снижении тарифа при увеличении данного параметра.

Для сравнения результатов для двух исследуемых предприятий были построены графики тарифов (рисунок 2) для $\alpha=0,1$. Тариф для металлургического предприятия выше по причине его рисковых характеристик и низкой эффективности предупредительных мероприятий по сравнению с электротехническим предприятием, являвшимся объектом исследования.

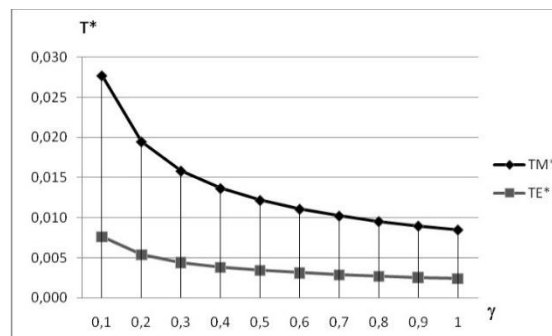


Рис. 2. Тарифы металлургического T_M^* и электротехнического T_E^* предприятий при $\alpha=0,1$

В результате решения задачи (1), (2), получены функциональные зависимости T^* , максимизирующие прибыль страховщика с учетом всех объектов, входящих в страховой портфель. Полученные функции позволяют определить оптимальные размеры тарифа по принимаемому на страхование риску с учетом параметром ущербов и выплат по застрахованным объектам, количество которых может превышать 10^7 (см. таблица 1).

В данном исследовании представлен результат только для портфеля по рисковым видам страхования, вероятность наступления страхового события которых описывается функцией $p = e^{-f\xi}$. В работе не учитываются законы распределения ущерба от застрахованного риска и распределение во времени в течение срока действия договора страхования.

Результаты могут быть использованы страховыми компаниями при формировании тарифной политики в области страхования промышленных рисков

ЛИТЕРАТУРА

- [1] Интернет-портал «Страхование сегодня» [Электронный ресурс]. – Режим доступа: <https://www.insur-info.ru>.
- [2] Лившиц, К.И. Оценка вероятности разорения страховой компании для модели Крамера-Лундберга с ММР-потокм страховых выплат / К.И. Лившиц // Новые информационные технологии в исследовании сложных структур. Материалы Двенадцатой конференции с международным участием. – 2018. – С. 128-129.
- [3] Береговая, М.П. Построение портфеля по модели У. Шарпа / М.П. Береговая // Разработка и применение наукоёмких технологий в целях модернизации современного общества. Сборник статей Всероссийской научно-практической конференции. – 2020. – С. 90-92.
- [4] Shi, P. Regression for copula-linked compound distributions with applications in modeling aggregate insurance claims / P. Shi, Z. Zhao // The Annals of Applied Statistics. – 2020. – Vol. 14(1). – P. 357-380.
- [5] Miljkovic, T. On two mixture-based clustering approaches used in modeling an insurance portfolio / T. Miljkovic, D. Fernández // Risks. – 2018. – Vol. 6(2). – P. 57.

Разработка методики сборки соплового аппарата с использованием методов кластеризации

В.А. Печенин
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
vadim.pechenin2015@gmail.com

Е.Ю. Печенина
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
ek-ko@list.ru

А.В. Куприянов
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
ssau.yap@gmail.com

Аннотация—Разработана методика для комплектации соплового аппарата, основанная на использовании численной модели, измеренных данных и методов кластеризации, направленная на снижение разноплощадности сечений между лопатками соплового аппарата. Проведены численные эксперименты по комплектации шести сопловых аппаратов с использованием разработанной методики комплектации. Разработанная методика показала лучшие результаты в сравнении с существующими подходами.

Ключевые слова— сборка, кластеризация, stl, сопловой аппарат, измеренные данные.

1. ВВЕДЕНИЕ

Актуальной задачей двигателестроения является снижение трудоемкости и повышение точности комплектации лопаточных машин. Лопатки ротора и статора изготавливаются с допусками, которые не позволяют обеспечивать полную взаимозаменяемость при сборке. В связи с этим лопатки комплектуют с дисками за несколько итераций – пересборок [1]. Традиционно комплектацию делают вручную с использованием специальных методик (сортировка по массе или по части отклонений геометрических параметров), что сопровождается пересборками всего комплекта [2], заменами лопаток и зачастую неоптимальностью достигаемых итоговых эксплуатационных параметров. Передовой технологией для контроля лопаток сопловых аппаратов является их 3D сканирование [3], и получаемые данные после соответствующей обработки можно использовать для имитационного моделирования сборки [4, 5] и численных расчетов оптимальных расстановок деталей в комплекте. В работе предложена методика, отличительной особенностью которой является использование цифровых моделей лопаток (по данным), компьютерного расчета оптимальной расстановки лопаток и методов кластеризации.

2. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Расстановка лопаток осуществляется по критерию разности разницу между наибольшей и наименьшей площадями проходных сечений (ППС) отдельных межлопаточных каналов в аппарате (разноплощадность Δ) и суммарный ППС. Авторами была разработана модель, описанная в [6], позволяющая на основании измеренных данных из файла *.stl (после сканирования) выполнять компьютерный расчет площади проходного сечения между двумя сопловыми лопатками. На рис. 1 приведена модель ППС между двумя лопатками.

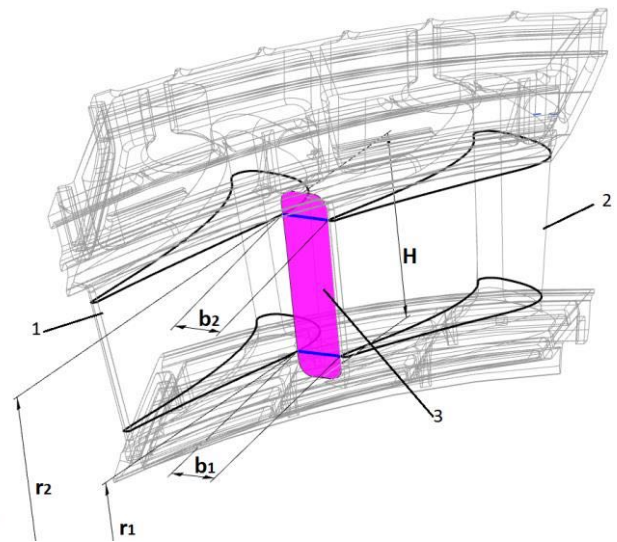


Рис. 1. Параметры площади: 1, 2 – первая и вторая лопатки, 3 – ППС; r_1 и r_2 – расстояния до сечений, b_1 и b_2 – кратчайшие расстояния в сечениях; H – расстояние между сечениями

Основная идея разработанной методики состоит в том, что рядом должны устанавливаться лопатки с близкими параметрами отклонений. Разработанная методика включает в себя четыре этапа: нормализация параметров (диапазон от 0 до 1); кластеризация на группы; сортировка центров кластеров; комплектация лопаток с использованием групп и модели расчета ППС. Разработанная методика применима как для расстановки лопаток внутри одного комплекта (имеются лопатки только для одного венца), так и для комплектации нескольких венцов. Для кластеризации использовался алгоритм k-means. Количество групп подбиралось исходя из того, что в классе не должно быть меньше пяти экземпляров и количество классов не меньше количества комплектов. На третьем этапе выполняется сортировка групп. Рассчитываются расстояния от центров (кластеров) каждой группы до нулевой точки. Так как параметры нормализованы в диапазон [0, 1], меньшее значение будет соответствовать 0, поэтому рассчитываются расстояния до нулевой точки в n-мерном пространстве нормированных параметров. Первым ставится группа с меньшими значениями координат центра, последней – с наибольшими (ближайшими к n-мерной точке из единиц).

На четвертом этапе выполняется собственно подбор лопаток. Из группы с наименьшими значениями координат центра выбирается лопатка, параметры которой ближе всего к нулевой точке. Лопатки располагаются по окружности. Затем выбирается лопатка

справа, параметры которой являются ближайшими к параметрам первой лопатки внутри первой группы. Рассчитывается ППС между этими двумя лопатками с помощью разработанной модели из [6]. Следующая лопатка, справа от первой, выбирается методом прямого подбора из оставшихся лопаток группы: рассчитываются ППС между первой лопаткой и остальными, выбирается ППС, имеющий меньшее отклонение от ППС между первой и второй лопатками. Четвертая лопатка подбирается уже справа от второй лопатке, аналогично как подбиралась третья, ППС сравнивается с ППС между третьей и первой лопаткой. Пятая ставится слева от третьей, шестая справа от четвертой. Когда заканчиваются лопатки первой группы, переходят к лопаткам второй группы и так далее. Когда комплект собран, следующий собирается по тем же шагам, описанным выше.

Было проведено имитационное моделирование лопаток для шести комплектов (пять параметров отклонений, 180 лопаток). Разработанная методика сравнивалась со случайной расстановкой и методикой, основанной на сортировке лопаток по возрастанию комплексного параметра (декартово расстояние из пяти рассматриваемых параметров).

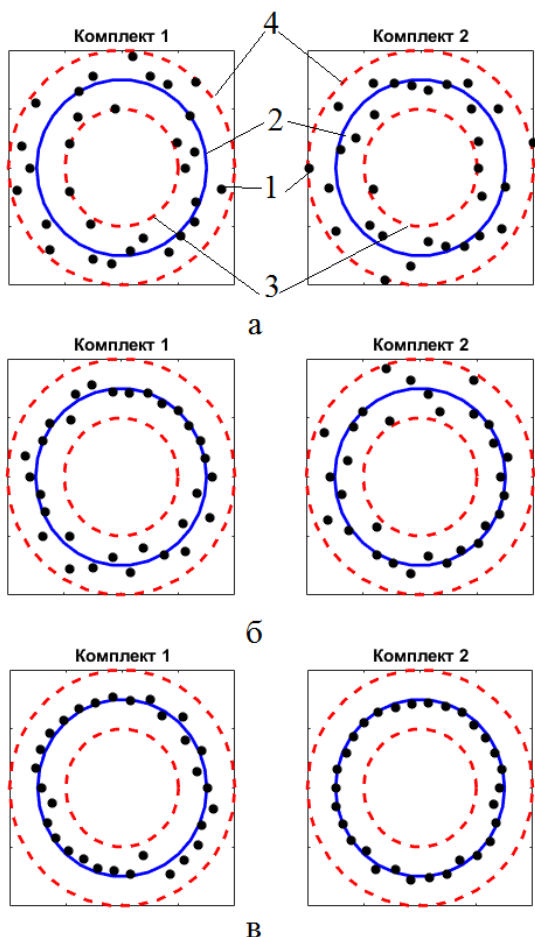


Рис. 2. Результаты ППС: 1 – значение для лопатки, 2 – конструкторское значение; 3, 4 – нижняя и верхняя границы поля допуска а) случайная расстановка б) методика с сортировкой; в) предлагаемая методика

Допуск на отклонение в горле решетки, согласно ОСТ 1 02571-86, составил $\pm 0,2$ мм. Суммарное номинальное значение ППС в сечении – 218,5686 мм. В алгоритме кластеризации использовалось 8 классов и метрика Евклида. Параметрами для кластеризации являлись величины отклонений формы и расположения каждой лопатки: смещения центров масс, отклонение угла закрутки и отклонение формы спинки и корыта. Величину Δ удалось снизить с 0,38–0,58 мм до 0,10 – 0,34 мм в шести комплектах. Отклонение в горле решетки не превысило допуск только при использовании разработанной методики.

3. ЗАКЛЮЧЕНИЕ

На основании проведенных вычислений выявлено, что изготовленные в допуске лопатки, отклонения формы и расположения которых подчиняются нормальному закону распределения, нельзя собирать по методу полной взаимозаменяемости, так как не соблюдаются требования по точности сборочного параметра, в частности величина ППС и разноплощадности Δ . Разработанную методику можно охарактеризовать как методику сборки, основанную на использовании метода индивидуального подбора с кластеризацией. Разработанная методика позволяет существенно улучшить достигаемые сборочные параметры с первой сборки.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках гранта Президента Российской Федерации (номер МК-2156.2022.1.6).

ЛИТЕРАТУРА

- [1] Олейникова, Е.В. Метод индивидуального подбора деталей как основа обеспечения качества сборки высокоточных изделий / Е.В. Олейникова, В.В. Непомилуев, А.Н. Семенов // Вестник Рыбинской государственной авиационной технологической академии им. П.А. Соловьева. – 2016. – № 4(39). – С. 37-42.
- [2] Осипович, Д.А. Исследование алгоритмов подбора лопаток при сборке сопловых аппаратов газотурбинного двигателя / Д.А. Осипович, С.Г. Ярушин, А.Б. Макеев // Сборка в машиностроении, приборостроении. – 2018. – № 7(216). – С. 313-319.
- [3] Расчёт фактической площади проходного сечения соплового аппарата (межлопаточных каналов лопаток) по данным оцифровки на измерительном комплексе ATOS II и программном обеспечении Geomagic Qualify // Методика №554-20-1. – ОАО "Авиадвигатель", 2011. – 63 с.
- [4] Ameta, G. Comparison of spatial math models for tolerance analysis: tolerance-maps, deviation domain, and TTRS / G. Ameta, S. Samper, M. Giordano // J Comput Inf Sci Eng. – 2011. – Vol. 11(2). – P. 255-267.
- [5] Das, A. Transfer function of assembly process with compliant non-ideal parts / A. Das, P. Franciosa, P.R.S. Prakash, D. Caglarelk // Procedia CIRP. – 2014. – Vol.21. – P. 177-182.
- [6] Pechenina, E.Yu. Model of analysis of the results of geometry scanning of turbine nozzles / E.Yu. Pechenina, V.A. Pechenin, M.A. Bolotov // IOP Conference Series: Materials Science and Engineering. – 2021. – Vol. 1155. – P. 012030.

Классификация по прецедентам и поиск в данных частых элементов

Н.А. Драгунов

Федеральный исследовательский
центр «Информатика и управление»
Российской академии наук
Москва, Россия
nikitadragunovjob@gmail.com

Е.В. Дюкова

Федеральный исследовательский
центр «Информатика и управление»
Российской академии наук
Москва, Россия
edjukova@mail.ru

А.П. Дюкова

Санкт-Петербургский научно-
исследовательский институт
эпидемиологии и микробиологии
им. Пастера
Санкт-Петербург, Россия
anastasia.d.95@gmail.com

Аннотация—Исследуются корректные логические классификаторы, при конструировании которых используются как основные идеи хорошо известного алгоритма «Кора», так и алгоритмов вычисления оценок. Предлагается модификация этапа обучения, позволяющая в определённых случаях сократить временные затраты без потери качества классификации. Приводятся результаты экспериментов на модельных и реальных данных и новые теоретические оценки сложности обучения классификаторов типа «Кора».

Ключевые слова— классификация по прецедентам, корректный логический классификатор типа «Кора», представительный элементарный классификатор, частый элементарный классификатор, тупиковое покрытие целочисленной матрицы.

1. ВВЕДЕНИЕ

В работе исследуются вопросы сокращения временных затрат на этапе обучения корректных классификаторов, базирующихся на применении методов дискретной математики. Одна из первых моделей классификаторов типа «Кора» предложена в [1]. В дальнейшем подход развивался в ряде работ ([2]– [4] и др.).

Обучение рассматриваемых классификаторов традиционно заключается в построении специальных фрагментов признаковых описаний прецедентов. Каждый такой фрагмент, называемый тупиковым представительным элементарным классификатором, позволяет отличать порождающий его объект от прецедентов из других классов, и при этом является в некотором смысле минимальным. Требование минимальности приводит к необходимости решать задачу построения тупиковых покрытий целочисленной матрицы, которая относится к числу труднорешаемых дискретных задач. Из найденных представительных элементарных классификаторов отбираются наиболее информативные, например те, которые достаточно часто встречаются в описаниях прецедентов. Вычисление оценки принадлежности распознаваемого объекта классу K осуществляется на основе проведения корректной процедуры голосования, в которой участвуют все отобранные представительные элементарные классификаторы класса K . Корректность голосования обеспечивает безошибочное распознавание прецедентов.

В настоящей работе предлагается модифицировать этап обучения на основе применения методов поиска частых (часто встречающихся) фрагментов в описаниях

прецедентов класса с последующим анализом встречаемости этих фрагментов в описаниях прецедентов из других классов.

Методы поиска частых элементов в данных обычно используются для построения ассоциативных правил. Наиболее востребованной областью их применения является анализ потребительской корзины [5].

Описываются результаты экспериментов на модельных и реальных данных. Приводятся теоретические выводы о сложности обучения классификаторов типа «Кора», базирующиеся на получении асимптотик для типичных значений наиболее важных количественных характеристик обучающей выборки.

2. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Рассматривается задача классификации по прецедентам с признаками x_1, \dots, x_n и непересекающимися классами K_1, \dots, K_l . Каждый признак имеет ограниченное число допустимых значений, которые кодируются целыми числами.

Пусть M – исследуемое множество объектов и пусть H – набор из r , $r \leq n$, различных признаков вида $\{x_{j_1}, \dots, x_{j_r}\}$; $\sigma = (\sigma_1, \dots, \sigma_r)$; σ_i – допустимое значение признака x_{j_i} , $i = 1, 2, \dots, r$. Пара (σ, H) называется элементарным классификатором (далее ЭК) ранга r .

ЭК (σ, H) порождает ЭК (σ', H') , если $\sigma' \subset \sigma$, $H' \subset H$. Объект S из M , имеющий признаковое описание (a_1, \dots, a_n) , содержит ЭК (σ, H) , если $a_{j_t} = \sigma_t$ при $t = 1, 2, \dots, r$.

Пусть K – некоторый класс объектов из M , $\bar{K} = \{K_1, \dots, K_l\} \setminus K$. Положим $Q(K)$ и $Q(\bar{K})$ – множество прецедентов соответственно из K и \bar{K} ; $|Q(K)| = m_1$, $|Q(\bar{K})| = m_2$, $1 \leq p \leq m_1$.

ЭК (σ, H) – p -частый в K , если не менее p объектов из $Q(K)$ содержат (σ, H) . ЭК (σ, H) называется максимальным p -частым в K , если (σ, H) – p -частый в K и в K не существует p -частого ЭК, порождающего (σ, H) .

ЭК (σ, H) ранга r называется правильным в K , если (σ, H) – r -частый в K .

ЭК (σ, H) – покрытие для \bar{K} , если ни один прецедент из $Q(\bar{K})$ не содержит (σ, H) . ЭК (σ, H) называется минимальным покрытием для \bar{K} , если (σ, H) – покрытие

для \bar{K} и (σ, H) не порождает никакого другого покрытия для \bar{K} .

ЭК (σ, H) – p -представительный для $Q(K)$, если (σ, H) – p -частый в $Q(K)$ и (σ, H) – покрытие для \bar{K} . ЭК (σ, H) называется *тупиковым* p -представительным для K , если (σ, H) – p -частый в K и (σ, H) – минимальное покрытие для \bar{K} .

В классическом варианте обучение классификатора рассматриваемого типа (модель A_1) основано на построении для каждого класса K тупиковых p -представительных ЭК. Для нахождения искомого ЭК перечисляются минимальные покрытия для \bar{K} , среди которых отбираются p -частые в K . В представляемой работе предложена и исследована модель классификатора A_2 , которая основана на перечислении максимальных p -частых ЭК класса K и отборе среди них тех, которые являются покрытиями для \bar{K} . В экспериментах для поиска частых ЭК описания прецедентов преобразовывались в бинарные вектора, далее применялся алгоритм FP-Growth [5]. На случайных данных показано, что новая модель классификатора (модель A_2) позволяют существенно сократить временные затраты при построении требуемых ЭК (см. рис. 1).

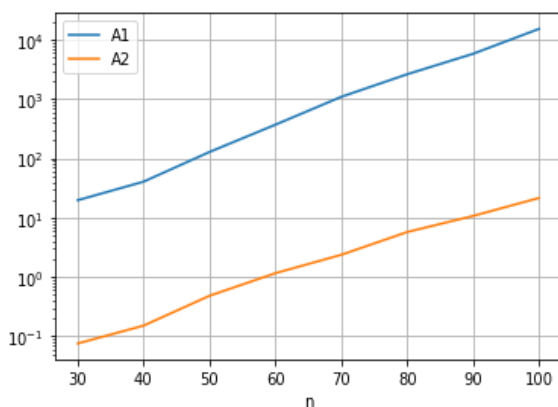


Рис. 1. Зависимость времени работы моделей A_1 и A_2 от n в секундах при $m_1 = 30$, $m_2 = 90$, $p = 1$ (логарифмическая шкала)

Результаты счёта на реальных целочисленных данных приведены в таблице 1 (k – максимальная значность признака). Две первые задачи “Молекулярная биология” и “Шахматы” взяты из репозитория UCI (<https://archive.ics.uci.edu/ml/datasets>). Третья задача предложена Федеральным медико-биологическим агентством. Модели A_1 и A_2 сравнивались по скорости работы и по точности классификации. В тестировании на точность также участвовал хорошо известный линейный классификатор Логистическая регрессия.

ТАБЛИЦА 1. ВРЕМЯ РАБОТЫ И ТОЧНОСТЬ НА ТЕСТОВОЙ ВЫБОРКЕ

Данные (m_1, m_2, n, k)	A_1 сек	A_1 точн.	A_2 сек	A_2 точн.	Лог. регр.
Задача 1 (767, 768, 61, 6)	3012.2	0.963	25.9	0.967	0.953
Задача 2 (1527, 1668, 73, 3)	323.6	0.951	180.7	0.980	0.965
Задача 3 (16, 63, 163, 2)	370.4	0.750	89.1	0.813	0.750

Представляет интерес получение асимптотических оценок для типичных значений количественных характеристик обучающей выборки, позволяющих оценить вычислительную сложность исследуемых моделей классификаторов в типичном случае. К таким характеристикам, в частности, относятся число частых ЭК класса K и число минимальных покрытий для \bar{K} , а также ранг ЭК указанного вида. В настоящей работе исследован случай бинарных данных.

В случае бинарных признаков множество $Q(K)$ представимо виде булевой матрицы L_K , имеющей размеры $m_1 \times n$. Введем обозначения: $M_{m_1 n}$ – множество всех булевых матриц размера $m_1 \times n$; $R(K)$ – число всех правильных ЭК в K ; $b_n \sim c_n$, $n \rightarrow \infty$, означает, что $\lim_{n \rightarrow \infty} b_n / c_n = 1$; $\phi(m_1)$ – интервал $(0.5 \log_2 m_1 n - 0.5 \log_2 \log_2 m_1 n - \log_2 \log_2 \log_2 n, 0.5 \log_2 m_1 n - 0.5 \log_2 \log_2 m_1 n + \log_2 \log_2 \log_2 n)$.

Теорема 1. Если $m_1^a \leq n \leq 2^{m_1}$, $a > 1$, то для почти всех матриц L_K из $M_{m_1 n}$ справедливо

$$R(K) \sim \sum_{r \in \phi(m_1)} C_n^r C_{m_1}^r 2^{r-r^2}, \quad n \rightarrow \infty,$$

и ранги почти всех правильных ЭК класса K принадлежат интервалу $\phi(m_1)$.

Аналогичные числовые характеристики для ЭК, являющихся минимальными покрытиями для \bar{K} , получены в [6]. Сравнение оценок из [6] с оценками, приведёнными в теореме 1, свидетельствует об эффективности в плане вычислительных затрат предлагаемого подхода к обучению классификаторов типа «Кора» в случае, когда m_2 не меньше m_1 , что согласуется с результатами экспериментов.

3. ЗАКЛЮЧЕНИЕ

Исследованы актуальные вопросы снижения временных затрат, возникающие при логическом анализе данных в задачах классификации на основе прецедентов. Теоретически и экспериментально обоснована целесообразность применения методов поиска частых элементов в данных на этапе обучения логического классификатора.

ЛИТЕРАТУРА

- [1] Баскакова, Л.В. Модель распознающих алгоритмов с представительными наборами и системами опорных множеств / Л.В. Баскакова, Ю.И. Журавлёв // Ж. вычисл. матем. и матем. физ. – 1981. – Т. 21, № 5. – С. 1264-1275.
- [2] Дюкова, Е.В. Поиск информативных фрагментов описаний объектов в дискретных процедурах распознавания / Е.В. Дюкова, Н.В. Песков // Ж. вычисл. матем. и матем. физ. – 2002. – Т. 42, № 5. – С. 711-723.
- [3] Дюкова, Е.В. О логическом анализе данных с частичными порядками в задаче классификации по прецедентам / Е.В. Дюкова, Г.О. Масляков, П.А. Прокофьев // Ж. вычисл. матем. матем. физики. – 2019. – Т. 59, № 9. – С. 1605-1616.
- [4] Kovshov, N.V. Algorithms for finding logical regularities in pattern recognition / N.V. Kovshov, V.L. Moiseev, V.V. Ryazanov // Computational Mathematics and Mathematical Physics. – 2008. – Vol. 48(2). – P. 314-328.
- [5] Aggarwal, Ch.C. Frequent Pattern Mining / Ch.C. Aggarwal, J. Han. – Springer International Publishing, 2014. – 471 p.
- [6] Дюкова, Е.В. Задача монотонной дуализации и её обобщения: асимптотические оценки числа решений / Е.В. Дюкова, Ю.И. Журавлёв // Ж. вычисл. матем. и матем. физ. – 2018. – Т. 58, № 12. – С. 2153-2168.

Метод комплексного анализа природных данных с применением искусственного интеллекта

В.В. Геппенер
Санкт-Петербургский
государственный

электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Санкт-Петербург, Россия
geppener@mail.ru

Б.С. Мандрикова
Институт космофизических
исследований и распространения
радиоволн ДВО РАН
Паратунка, Камчатский край, Россия
555bs5@mail.ru

Н.Г. Пятайчук
Санкт-Петербургский
государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Санкт-Петербург, Россия
pet.ai.4.uk@yandex.ru

Аннотация—Предложен метод комплексного анализа сложно структурированных природных данных, основанный на комбинации нейронных сетей с конструкциями вейвлет-преобразования. На примере данных вторичных космических лучей, подтверждена эффективность метода для детектирования и оценки параметров аномальных вариаций. Эмпирически доказано, что совмещение нейронных сетей Автокодировщик и пакетных вейвлет-разложений позволяет извлекать сложные зависимости в данных и подавлять шум.

Ключевые слова— данные сложной структуры, нейронные сети, вейвлет-преобразование, космические лучи.

1. ВВЕДЕНИЕ

Со стремительным развитием техносферы, основанном на прогрессе науки и технологий, все сильнее ощущается неудовлетворительность существующих методов обработки и анализа сложно структурированных данных. Особая сложность возникает при исследовании природных данных ввиду отсутствия знаний об информативной составляющей сигнала и, как следствие, невозможности эффективного вычищения шума различной природы. Примером таких исследований являются факторы в данных геофизического мониторинга при выполнении прогноза космической погоды. Аномальные события в околоземном космическом пространстве способны создать радиационную опасность для пилотов высотной авиации и космонавтов, сбой радиосвязи, а также могут привести к потере ориентации и разрушению спутников [1]. Космические лучи являются одним из значимых факторов космической погоды. В настоящее время исследованием космических лучей занимается большое число ученых и научных групп, однако вопрос оперативного обнаружения аномальных вариаций с требуемой точностью остается открытым [1].

В докладе представлен метод комплексного анализа данных вторичных космических лучей и детектирования аномальных вариаций на основе применения нейронной сети (НС) Автокодировщик, пакетных вейвлет-разложений и непрерывного вейвлет-преобразования. Сеть Автокодировщик эффективна для подавления шума и извлечения зависимостей в сложных данных [2]. Вейвлет-преобразование позволяет детально изучить частотно-временную структуру сигнала, выделить информативные компоненты и подавить шум [3]. Данная работа является продолжением исследования [4]. Для снижения риска наступления ложной тревоги предложено правило, дающее результат по совокупности решений.

2. ОПИСАНИЕ МЕТОДА

Предлагаемый метод комплексного анализа сложно структурированных данных включает следующие операции:

1. Преобразование данных КЛ на основе Автокодировщика:

$$\hat{f} = h^{(2)}(\omega^{(2)}(h^{(1)}(\omega^{(1)}F + b^{(1)})) + b^{(2)}), \quad (1)$$

где $h^{(1)}$ передаточная функция энкодера, $\omega^{(1)}$, $\omega^{(2)}$ матрицы весов, $h^{(2)}$ – передаточная функция декодера, $b^{(1)}$, $b^{(2)}$ – векторы смещения.

2. Применение непрерывного вейвлет-преобразования: $W\hat{f}(u, s) = \int_{-\infty}^{+\infty} \hat{f}(t) \frac{1}{\sqrt{s}} \Psi^*\left(\frac{t-u}{s}\right) dt$, где Ψ – вейвлет, s – масштаб, u – сдвиг по времени, $s, u \in R$ (R – действительные числа) $s \neq 0$.

3. Применение пороговой функции:

$$P_{T_s^l, NN} [W\hat{f}(u, s)] = \begin{cases} W\hat{f}(u, s), & \text{если } |W\hat{f}(u, s)| \geq T_s^l, \\ 0, & \text{если } |W\hat{f}(u, s)| < T_s^l, \end{cases} \quad (2)$$

где $T_s^l = q \times \sigma_s^l$, σ_s^l – среднеквадратическое отклонение (СКО) коэффициентов, рассчитанное в скользящем окне длины l , q – пороговый коэффициент.

4. Оценка интенсивности аномалий в момент времени $t=u$:

$$E(u)_{NN} = \sum_s P_{T_s^l, NN} [W\hat{f}(u, s)]. \quad (3)$$

5. Выполняем разложение функции f в вейвлет-пакеты: $W_j^0 = \bigoplus_{i=0}^l W_{j_i}^{P_i}$, $\{\Psi_{j_i}^{P_i}(2^{j_i}t - m)\}_{m \in N}$ есть базис пространства $W_{j_i}^{P_i}$;

6. Определяем ветви дерева, соответствующие компонентам сигнала, базис $O_{j_i}^p$ пространства $W_{j_i}^{P_i}$ есть базис:

$$O_{j_i}^p = \begin{cases} \{\Psi_{j_i}^{P_i}(2^{j_i}t - m)\}_{m \in Z}, \\ \sum_{m \in I_M^p} |\langle X, \Psi_{j_i, m}^p \rangle|^2 \geq \sum_{m \in I_M^{2p}} |\langle X, \Psi_{j_i+1, m}^{2p} \rangle|^2 + \sum_{m \in I_M^{2p+1}} |\langle X, \Psi_{j_i+1, m}^{2p+1} \rangle|^2, \\ \{\Psi_{j_i}^{2p}\}_{m \in Z} \cup \{\Psi_{j_i+1}^{2p+1}\}_{m \in Z}, \\ \sum_{m \in I_M^p} |\langle X, \Psi_{j_i, m}^p \rangle|^2 < \sum_{m \in I_M^{2p}} |\langle X, \Psi_{j_i+1, m}^{2p} \rangle|^2 + \sum_{m \in I_M^{2p+1}} |\langle X, \Psi_{j_i+1, m}^{2p+1} \rangle|^2, \end{cases}$$

где множество индексов I_M^l , $l = P, 2P, 2P+1$ определяются так: индекс $m \in I_M^l$, если $|\langle X, \Psi_{j_i, m}^l \rangle| \geq T_{j_i}$ порог $T_{j_i} = K * \sigma_{j_i}^l$,

$$\sigma_{j_i}^l = \sqrt{\frac{1}{l} \sum_{m=1}^l \left(\langle X, \Psi_{j_i, m}^l \rangle - \langle X, \Psi_{j_i, m}^l \rangle \right)^2}.$$

Получаем представление сигнала в виде:

$$\tilde{f}(t) = \sum_{(j_i, m) \in Q_M^l} \langle X, \Psi_{j_i, m}^l \rangle \Psi_{j_i, m}^l(t) = \sum_{(j_i, m) \in Q_M^l} v_{j_i, m} \Psi_{j_i, m}^l(t), \quad (4)$$

где $(j_i, m) \in Q_M^l : j_i \in O_M^p, m \in I_M^l$.

7. Применение непрерывного вейвлет-преобразования:

$$W\hat{f}(u, s) = \int_{-\infty}^{+\infty} \hat{f}(t) \frac{1}{\sqrt{s}} \Psi^* \left(\frac{t-u}{s} \right) dt, \quad (4)$$

где Ψ – вейвлет, s – масштаб, u – сдвиг по времени, $s, u \in R$ (R – действительные числа) $s \neq 0$.

8. Применение пороговой функции:

$$P_{T_s^l}^{pac} [W\hat{f}(u, s)] = \begin{cases} W\hat{f}(u, s), & \text{если } |W\hat{f}(u, s)| \geq T_s^l, \\ 0, & \text{если } |W\hat{f}(u, s)| < T_s^l, \end{cases} \quad (5)$$

где $T_s^l = q \times \sigma_s^l$, σ_s^l – средноквадратическое отклонение (СКО) коэффициентов, рассчитанное в скользящем окне длины l , q – пороговый коэффициент.

9. Оценка интенсивности аномалий в момент времени $t=u$:

$$E(u)_{pac} = \sum_s P_{T_s^l}^{pac} [W\hat{f}(u, s)]. \quad (6)$$

10. Аномалия в данных есть, если выполняются условия: $E(u)_{NN} > T_{аном}$ и $E(u)_{pac} > T_{аном}$ при $T_{аном} = 1.5$.

На Рис. 1 представлен результат применения предлагаемого метода к минутным данным нейтронного монитора ст. Оулу. Оранжевыми вертикальными линиями отмечены моменты регистрации геомагнитных бурь. На Рис. 1б,в представлен результат применения операций (2) и (3), соответственно, к исходным данным нейтронного монитора. На Рис. 1д,е представлен результат применения операций (2) и (3) к данным нейтронного монитора после применения нейронной сети. На Рис. 1з,и представлен результат операций (5) и (6) к данным нейтронного монитора после вейвлет-фильтрации по формуле (4). Результаты показывают, что предобработка данных на основе нейронной сети (операция (1)) (Рис. 1з,д) и выделение когерентных структур с использованием вейвлет-пакетов (операция (4)) (Рис. 1з,и) позволяют существенно снизить уровень шума и выделить аномальные вариации – аномальные понижения и повышения интенсивности космических лучей.

Результат показывает эффективность предлагаемого метода для анализа вторичных космических лучей и выделения аномальных событий в космических лучах в период магнитных бурь.

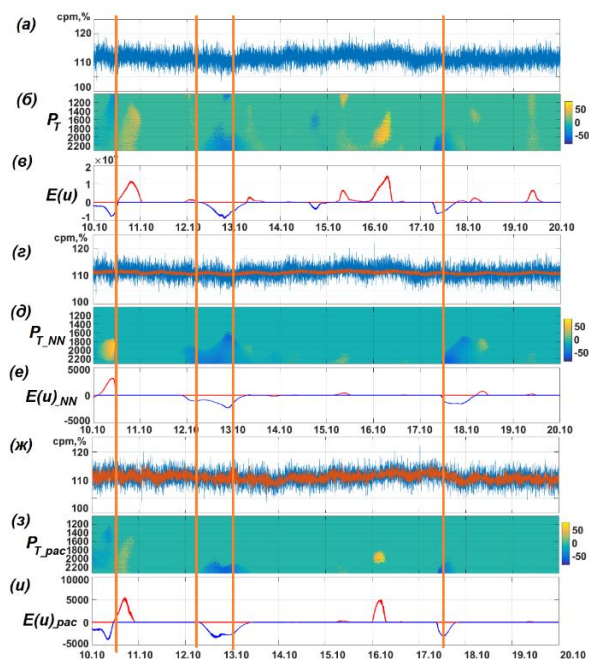


Рис. 1. Результат применения метода к данным НМ ст. Оулу за 10-20 октября 2021 г.: а) данные НМ; б) результат оп. (2) к данным НМ; в) результат оп. (3) к данным НМ; г) синим цветом изображены данные НМ, оранжевым – результат оп. (1); д) результат оп. (3) к данным НМ после аппроксимации НС; е) результат оп. (3) к данным НМ после аппроксимации НС; ж) синим цветом изображены данные НМ, оранжевым – результат оп. (4); з) результат оп. (5) к данным НМ после вейвлет-фильтрации по формуле (4); и) результат оп. (6) к данным НМ после вейвлет-фильтрации

3. ЗАКЛЮЧЕНИЕ

Предложенный метод комплексного анализа сложно структурированных природных данных показал свою эффективность для выделения аномальных особенностей различной интенсивности и продолжительности. Апробация метода с использованием данных вторичных космических лучей подтвердила его результативность для детектирования аномальных событий в сигнале в периоды магнитных бурь. Результаты показали, что совмещение нейронной сети Автокодировщик с пакетными вейвлет-разложениями позволяет извлекать сложные зависимости в данных и подавлять шум.

БЛАГОДАРНОСТИ

Работа выполнена в рамках Государственного задания по теме “Физические процессы в системе ближнего космоса и геосфер при солнечных и литосферных воздействиях” (2021–2023 гг.), регистрационный номер АААА-А21-121011290003-0.

ЛИТЕРАТУРА

- [1] Кузнецов, В.Д. Космическая погода и риски космической деятельности / В.Д. Кузнецов, // Космическая техника и технологии. – 2014. – № 3(6). – С. 3-13.
- [2] Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль. – «ДМК Пресс», 2018. – 652 с.
- [3] Малла, С. Вейвлеты в обработке сигналов / С. Малла. – «Мир», 2005. – 671 с.
- [4] Geppener, V. Combination of wavelet transform and Autoencoder for complex data analysis and anomaly detection / V. Geppener, B. Mandrikova // International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649436.

Анализ очередей для видеотрафика с помощью обобщённого интервального метода

Б.Я. Лихтциндер
Поволжский государственный
университет телекоммуникаций и
информатики
Самара, Россия
lixt@psuti.ru

Е.В. Китаева
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
el_kitaeva@mail.ru

А.Ю. Привалов
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
privalov1967@gmail.com

Аннотация—В докладе представлен новый подход к оценке средней очереди в системах передачи данных, трафик в которых очень коррелирован и практически не описывается классическими моделям теории очередей. В частности, такая ситуация имеет место в системах передачи видеотрафика. Предлагается подход, в котором модель системы передачи данных описывается не двумя случайными процессами – входным потоком и процессом обслуживания, а одним комбинированным процессом. Для вывода обобщённой формулы для средней очереди при этом используется обобщённый интервальный метод. Получена явная зависимость средней очереди от корреляционных характеристик комбинированного процесса.

Ключевые слова— Моделирование трафика, буферизация видеотрафика, интервальный метод анализа очередей.

1. ВВЕДЕНИЕ

При традиционном подходе к описанию систем передачи данных с помощью моделей теории массового обслуживания, главными характеристиками её являются два случайных процесса – входной поток требований в систему и процесс обслуживания поступивших требований. Однако, в современных сетях телекоммуникаций, передающих сильно коррелированный пачечный трафик (такой, как видеотрафик), классические модели таких процессов являются мало адекватными.

Для преодоления этих трудностей в данном докладе предлагается использовать подход, когда система массового обслуживания описывается одним случайным процессом – количеством заявок, приходящих за время обслуживания одной заявки.[2] В качестве примера применения подобного подхода рассматривается вывод формулы для средней очереди в системе массового обслуживания, на вход которой подаётся реальный видеотрафик. При этом используется обобщённый интервальный метод, с помощью которого удаётся получить выражение для средней очереди через корреляционные характеристики указанного выше процесса. Данный подход является обобщением подхода, предложенного в [1].

Корректность полученных результатов проверяется с помощью имитационного моделирования.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассматривается одноканальная система массового обслуживания с бесконечной очередью, входным потоком общего вида без ограничений на

корреляционные свойства, и временем обслуживания общего вида.

Для описания работы системы построим следующий случайный процесс: пусть $\theta_k = t_{k+1} - t_k$, где t_k – время ухода из системы k -той заявки, то есть, θ_k – это интервал между моментами окончания обслуживания последовательных заявок. Пусть $m(\theta_k)$ – количество заявок, приходящих в систему за интервал θ_k , а $q(\theta_k)$ – очередь в системе непосредственно перед окончанием интервала θ_k . Тогда очередь в системе можно описать уравнением

$$q(\theta_k) = q(\theta_{k-1}) + m(\theta_k) - \delta(\theta_k) \quad (1)$$

где $\delta(\theta_k) = 0$, если $q(\theta_{k-1}) = m(\theta_k) = 0$ и $\delta(\theta_k) = 1$ в противном случае.

Циклом обслуживания будем называть совокупность смежных интервалов $\theta_k, \dots, \theta_{k+N-1}$, такую, что $q(\theta_{k-1}) = q(\theta_{k+N-1}) = 0$, а для всех остальных θ_i , входящих в этот цикл, $q(\theta_i) > 0$. Данное определение является обобщением определения, введённого в [1].

Обратим внимание на отличие известного понятия периода занятости, широко используемого в теории массового обслуживания, от введённого понятие цикла обслуживания. Период занятости – это интервал времени между моментами, когда обслуживающий прибор свободен, а цикл обслуживания – это период времени между моментами, когда очередь равна 0, но прибор занят. Таким образом, в одном периоде занятости может быть несколько циклов обслуживания. Циклы обслуживания плотно занимают весь интервал, где прибор занят, рассматривая поведение системы в моменты θ_k равносильно исключению из рассмотрения интервалов, когда прибор не работает.

Введём гипотезу эргодичности, аналогично тому, как это было сделано в [1]. Мы будем предполагать, что случайные процессы $m(\theta_k)$ и $q(\theta_k)$ стационарны, взаимно стационарны и эргодичны относительно статистических характеристик первого и второго порядка, в том числе и взаимных корреляционных моментов. Далее везде будем считать, что гипотеза справедлива.

Возведём в квадрат обе части (1) и учтём, что

$$\delta^2(\theta_k) = \delta(\theta_k), \quad \delta(\theta_k) m(\theta_k) = m(\theta_k), \quad \delta(\theta_k) q(\theta_{k-1}) = q(\theta_{k-1}).$$

Получим

$$q^2(\theta_k) = q^2(\theta_{k-1}) + 2 q(\theta_{k-1}) m(\theta_k) - 2 q(\theta_{k-1}) + m^2(\theta_k) - 2m(\theta_k) + \delta(\theta_k).$$

Вычислим математическое ожидание от обеих частей этого равенства, учитывая, что из условия стационарности следует, что

$$M(\delta(\theta_k)) = M(m(\theta_k)), \quad M(q^2(\theta_k)) = M(q^2(\theta_{k-1})),$$

получим

$$M(q(\theta_{k-1})) = 1/2 * (M(m^2(\theta_k)) - M(m(\theta_k))) + M(q(\theta_{k-1})m(\theta_k)).$$

Рассуждая аналогично [1], $M(q(\theta_{k-1})m(\theta_k))$ можно выразить через эмпирические корреляционные характеристики процесса $m(\theta_k)$:

$$v(M, k) = \frac{1}{M-k} \sum_{i=k+1}^M m(\theta_i)(m(\theta_{i-k}) - 1)$$

$$M(q(\theta_i)m(\theta_j)) = \lim_{M \rightarrow \infty} \sum_{k=1}^{M-1} \frac{M-k}{M} v(M, k)$$

где использовалось равенство

$$q(\theta_{i-1}) = \sum_{k=j_s}^{i-1} (m(\theta_k) - 1)$$

где j_s – первый интервал θ_k в том цикле обслуживания, в котором находится и θ_k , а само равенство следует из того, что внутри цикла обслуживания прибор не простаивает (детальное доказательство см. в [1]).

Кроме того, внутри цикла обслуживания $M(\delta(\theta_k))=1$, а потому и $M(m(\theta_k))=1$, и $M(m^2(\theta_k)) - M(m(\theta_k)) = D(m(\theta_k))$, то есть дисперсии $m(\theta_k)$. Собирая всё вместе, получаем

$$M(q(\theta_i)) = \frac{1}{2} D(m(\theta_k)) + \lim_{M \rightarrow \infty} \sum_{k=1}^{M-1} \frac{M-k}{M} v(M, k)$$

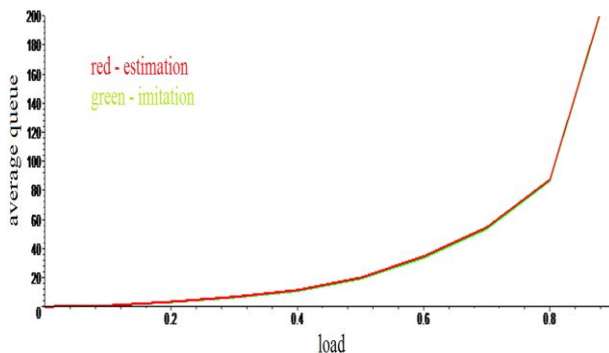


Рис.1. Средняя очередь для видеотрафика, θ_k пропорциональны размерам пакетов

Заметим, что мы нашли фактически математическое ожидание очереди при условии, что прибор занят. Чтобы найти безусловное математическое ожидание очереди, вспомним, что доля времени, когда прибор занят, составляет $1-\rho$, где ρ – нагрузка прибора, и на интервалах

простоя прибора очередь, очевидно, равна нулю, поэтому окончательная формула для средней очереди:

$$\bar{q} = \rho \left(\frac{D(m(\theta_k))}{2} + \lim_{M \rightarrow \infty} \sum_{k=1}^{M-1} \frac{M-k}{M} v(M, k) \right)$$

В докладе будут представлены результаты расчётов по данной формуле, проведённые для трасс реального видеотрафика в сравнении с результатами имитационного моделирования прохождения того же видеотрафика через соответствующую систему массового обслуживания. На рис.1 приведён пример такого сравнения.

3. ЗАКЛЮЧЕНИЕ

Представлена обобщённая формула средней очереди для входного трафика с произвольными корреляционными свойствами. В формуле присутствуют только моменты второго порядка комбинированного случайного процесса, описывающего одновременно входной поток и процесс обслуживания, и загрузка канала передачи данных. Полученная формула хорошо согласуется с данными, полученными на реальном видеотрафике.

ЛИТЕРАТУРА

- [1] Лихтциндер, Б.Я. Об оценках средней длины очереди для одноканальных систем массового обслуживания через статистические безусловные моменты второго порядка модифицированного входного потока / Б.Я. Лихтциндер, И.А. Блатов, Е.В. Китаева // Автоматика и телемеханика. – 2022. – № 1. – С. 113-129.
- [2] Лихтциндер, Б.Я. Трафик мультисервисных сетей доступа (анализ и проектирование) / Б.Я. Лихтциндер.- М.: Горячая линия – Телеком, 2018. – 200 с.
- [3] Likhttsinder, B. Traffic Buffering in Multichannel Switches of Access Networks / B. Likhttsinder, A. Privalov // International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649310.
- [4] Leland, W.E. On the self-similar nature of Ethernet traffic / W.E. Leland, S. Taqqu Murad, W. Willinger, D.V. Wilson // J. IEEE/ACM Trans. Networking. – 1994. – Vol. 2(1). – P. 1-15.
- [5] Neuts, M.F. Versatile Markovian point process / M.F. Neuts // J. Appl. Probab. – 1979. – Vol. 16(4). – P. 764-779.
- [6] Ramaswami, V. The N/G/1 queue and its detailed analysis / V. Ramaswami // Adv. Appl. Probab. – 1980. – Vol. 12(1). – P. 222-261.
- [7] Jagerman, D.L. Mean waiting time approximations in the G/G/1 queue / D.L. Jagerman, B. Balcioglu, T. Altioik, B. Melamed // Queueing Syst. – 2004. – Vol. 46. – P. 481-506.
- [8] Balcioglu, B. Approximate mean waiting time in a GI/D/1 queue with autocorrelated time to failures / B. Balcioglu, D.L. Jagerman, T. Altioik // IEEE Trans. – 2007. – Vol. 39(10). – P. 985-996.
- [9] Kartashevskii, I.V. Processing of correlated traffic at a G/G/1 type node / I.V. Kartashevskii, A.V. Saprykin // Radiotekhnika. – 2017. – Vol. 10. – P. 119-125.
- [10] Tsybakov, B.S. Teletraffic model based on a self-similar random process / B.S. Tsybakov // Radiotekhnika. – 1999. – Vol. 5. – P. 24-31.
- [11] Shelukhin, O.I. Fractal Processes in Telecommunications / O.I. Shelukhin, A.M. Tenyakshev, A.V. Osin. – Moscow: Radiotekhnika, 2003.

Оценка параметров движения объекта в условиях изменения качества измерительных данных

Ю.В. Цыганова
Ульяновский государственный
университет
Ульяновск, Россия
tsyganovajv@gmail.com

А.А. Калянов
Ульяновский государственный
университет
Ульяновск, Россия
andre756@yandex.ru

О.В. Лукин
Ульяновский государственный
университет
Ульяновск, Россия
oleg.lukin.v@mail.ru

Аннотация—Работа посвящена решению задачи оценивания параметров движения объекта при непредвиденном изменении качества измерительных данных. Рассматриваются три модели движения: статический объект, прямолинейное движение и круговое движение. Решение задачи получено в форме адаптивного расширенного фильтра Калмана с динамической настройкой ковариационной матрицы шума измерений. Предложенный алгоритм реализован и протестирован на языке MATLAB.

Ключевые слова— Модель движения объекта, данные измерений, ковариация шума измерений, расширенный фильтр Калмана, адаптивная фильтрация.

1. ВВЕДЕНИЕ

В настоящее время задача оценивания параметров движения объекта является важной для многих практических приложений [1]. Рассмотрим актуальную задачу оценивания параметров движения объекта на плоскости по данным измерений, получаемых от конечного множества датчиков. Датчики расположены на плоскости, их координаты считаем известными.

Предположим, что траекторию движения объекта можно представить дискретной стохастической моделью с вектором состояния $X(t_k)$, где t_k – дискретный момент времени, $k=1,2,\dots$. Параметрами движения объекта считаем элементы вектора $X(t_k)$, например, координаты на плоскости Ox , проекции скорости и т. д. В каждый дискретный момент времени измеряют расстояние от объекта до каждого датчика. Моделирование и оценивание вектора $X(t_k)$ по данным измерений можно выполнить с помощью известного фильтра Калмана.

В [2] рассмотрено решение данной задачи при условии, что значение ковариационной матрицы шума измерений совпадает с фактической ковариационной матрицей. Однако в реальных условиях ковариация шума измерений может возрасти на каком-то участке движения объекта (например, при наличии препятствия для передачи сигнала с датчиков). Тогда ошибки оценок, вычисляемые в алгоритме фильтрации, будут возрастать. Для того чтобы решить указанную проблему, применяют методы, основанные на адаптивном оценивании величины ковариационной матрицы шума измерений (краткий обзор методов см., например, в [3]).

Целью данной работы является построение практического алгоритма для оценки параметров движения объекта в условиях непредвиденного увеличения уровня помех в измерительных данных.

2. АЛГОРИТМ ОЦЕНКИ ПАРАМЕТРОВ ДВИЖЕНИЯ ОБЪЕКТА

Рассмотрим три модели движения на плоскости Ox : статический объект, прямолинейное движение, круговое движение. Эти виды движения характерны для мобильных робототехнических устройств.

Предположим, что каждый из датчиков, с которых поступают измерительные данные о текущем расстоянии до объекта, имеет погрешность, которая моделируется гауссовской случайной величиной с нулевым средним и дисперсией σ_m^2 . При движении объекта в некоторый дискретный момент времени дисперсия шума может возрасти до уровня $\mu \cdot \sigma_m^2$. На рис. 1 и 2 показан пример прямолинейного движения и данных с датчика 1.

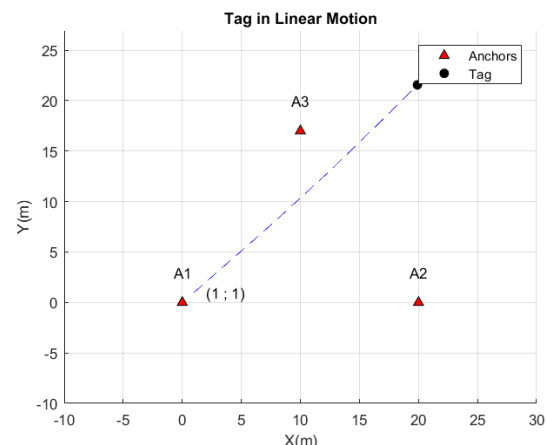


Рис. 1. Результат имитационного моделирования траектории движения объекта (равномерное прямолинейное движение)

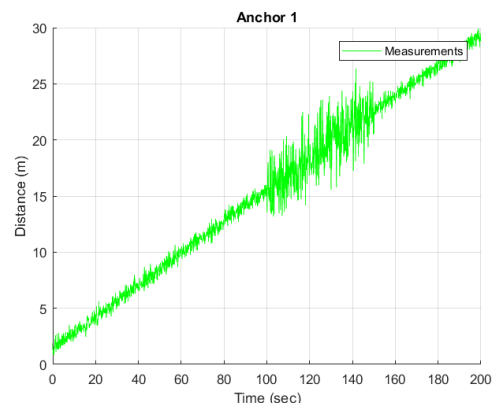


Рис. 2. Результат имитационного моделирования измерений с первого датчика (равномерное прямолинейное движение)

На рис. 3 и 4 показан пример кругового движения и данных с датчика 1.

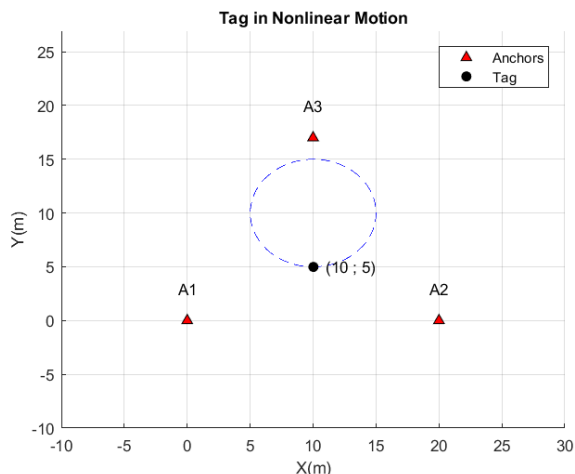


Рис. 3. Результат имитационного моделирования траектории движения объекта (круговое равномерное движение)

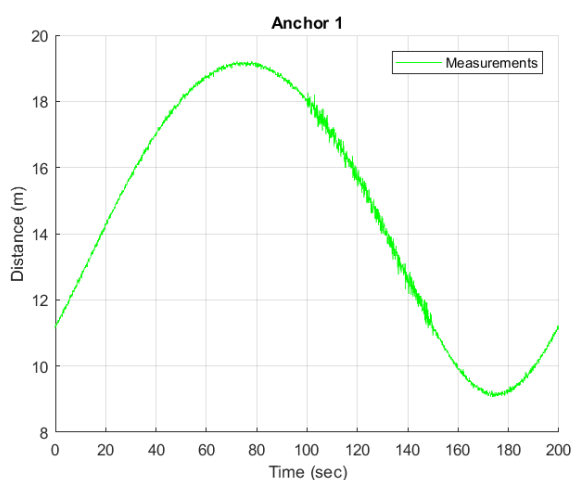


Рис. 4. Результат имитационного моделирования измерений с первого датчика (круговое равномерное движение)

На рис. 2 и 4 можно видеть, что на временном интервале от 100 до 150 сек погрешность измерений возрастает.

В настоящей работе для решения указанной задачи построен алгоритм оценивания вектора состояния модели движения объекта в форме адаптивного расширенного фильтра Калмана с динамической настройкой значения ковариационной матрицы ошибок измерений.

Алгоритм основан на модификации решения, предложенного в [4] для линейной стохастической модели сигнала и линейного дискретного фильтра. В нашем случае модель измерений нелинейная, поэтому мы применяем вместо стандартного фильтра Калмана расширенный фильтр.

Алгоритм адаптивной фильтрации с динамической оценкой матрицы ковариации шума измерений состоит из нескольких этапов, включая:

- 1) вычисление оценок вектора состояния и ковариационных матриц ошибок оценивания на этапах предсказания и фильтрации,
- 2) вычисление расчетной и фактической ковариационной матрицы ошибок измерений,
- 3) вычисление специальной сигнальной функции,
- 4) коррекция матрицы Калмана в уравнениях фильтра,
- 5) вычисление оценки ковариационной матрицы шума измерений.

Результаты имитационного моделирования в среде MATLAB для трех моделей движения подтверждают работоспособность предложенного решения.

3. ЗАКЛЮЧЕНИЕ

В работе построен и реализован алгоритм оценивания параметров движения объекта при непредвиденном увеличении уровня помех в измерительных данных. Алгоритм является адаптивным, поскольку позволяет динамически настроить величину ковариационной матрицы ошибок измерений. Тестирование проведено для трех моделей движения: статический объект, прямолинейное движение и круговое движение.

Результаты проведенных вычислительных экспериментов подтверждают, что при применении предложенного алгоритма величина ошибки оценки вектора состояния модели объекта уменьшается по сравнению с обычным (не адаптивным) фильтром. Результаты работы могут быть использованы при разработке программного обеспечения для мобильных робототехнических устройств.

ЛИТЕРАТУРА

- [1] Bar-Shalom, Y. Estimation with Applications to Tracking and Navigation / Y. Bar-Shalom, T. Kirubarajan, X.-R. Li. – John Wiley & Sons, Inc., 2002.
- [2] Семушин, И.В. Моделирование и оценивание траектории движущегося объекта / И.В. Семушин, А.В. Цыганов, Ю.В. Цыганова, А.В. Голубков, С.Д. Винокуров // Вестник ЮУрГУ. Серия “Математическое моделирование и программирование”. – 2017. – Т. 10, № 3. – С. 108-119. DOI: 10.14529/mmp170309.
- [3] Chhabra, A. Measurement Noise Covariance-Adapting Kalman Filters for Varying Sensor Noise Situations / A. Chhabra, J.R. Venepally, D. Kim // Sensors. – 2021. – Vol. 21. – P. 8304. DOI: 10.3390/s21248304.
- [4] Semushin, I.V. Stable method of estimation of parameters of a linear filter / I.V. Semushin, A.G. Skovikov, L.V. Kalinin, Yu.V. Tsyganova // Measurement Techniques. – 1999. – Vol. 42(9). – P. 848-852. DOI: 10.1007/BF02504263.

Параметрическая идентификация распределенной мультисенсорной системы фильтрации

И.В. Семушин
Ульяновский государственный
университет
Ульяновск, Россия
kentvsem@gmail.com

Ю.В. Цыганова
Ульяновский государственный
университет
Ульяновск, Россия
tsyganovajv@gmail.com

А.В. Цыганов
Ульяновский государственный
педагогический университет
им. И.Н. Ульянова
Ульяновск, Россия
andrew.tsyganov@gmail.com

Аннотация—Работа посвящена решению задачи параметрической идентификации распределенной мультисенсорной системы фильтрации. Для решения задачи предложен новый гибридный алгоритм параметрической идентификации, основанный на применении метода вспомогательного функционала качества (ВФК). Впервые показано, как можно вычислить градиент критерия ВФК в распределенной мультисенсорной системе. Алгоритм реализован и протестирован в программной среде MATLAB.

Ключевые слова —параметрическая идентификация, мультисенсорная система, распределенная фильтрация Калмана, вспомогательный функционал качества, гибридный алгоритм.

1. ВВЕДЕНИЕ

В настоящее время одним из наиболее популярных подходов к фильтрации измерительных данных являются методы фильтрации Калмана, которые применяются в различных областях исследований и во многих практических приложениях [1].

Различают два типа калмановской фильтрации: ковариационная и информационная. Если ковариационная форма фильтра Калмана уже давно является “стандартным” алгоритмом, то информационная форма приобрела известность при решении задач распределенной фильтрации в мультисенсорных динамических системах [2].

Целью данной работы является решение вопроса о применимости метода вспомогательного функционала качества (ВФК) в задаче параметрической идентификации распределенной мультисенсорной системы фильтрации. Предложенное новое решение основано на информационном фильтре Калмана.

2. МЕТОД ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Рассмотрим распределенную мультисенсорную систему фильтрации полезного сигнала $x(t_k)$ по данным измерений $y_i(t_k)$, доступным от $i=1,2,\dots,m$ сенсоров в дискретные моменты времени $t_k=kT$ ($k=1,2,\dots,N$), где T обозначает заданный темп измерений. Пусть математическая модель данной системы представлена разностными уравнениями [3]:

$$\left. \begin{aligned} x(t_k) &= dx(t_{k-1}) + A\sqrt{1-d^2}w(t_k) \\ y_i(t_k) &= x(t_k) + v_i(t_k), \quad i = 1, 2, \dots, m \end{aligned} \right\} \quad (1)$$

где $w(t_k)$ и $v_i(t_k)$ – взаимно независимые дискретные белые шумы с нулевым средним и дисперсиями $Q=1$ и $R_i=B_i^2$, соответственно.

Обозначим через $\theta=(d, A^2, B_1^2, \dots, B_m^2)^T$ – параметры модели (1). Если значение θ известно, тогда фильтрация полезного сигнала $x(t_k)$ по данным мультисенсоров $y_i(t_k)$ может быть выполнена с помощью информационного фильтра Калмана. В данной работе мы рассматриваем систему (1) с параметрической неопределенностью, т. е. в предположении, что значение параметра θ не известно.

Задача параметрической идентификации заключается в вычислении оценок параметра неопределенности θ по данным мультисенсоров $y_i(t_k)$. Для решения применим метод ВФК [4]. Известно, что данный метод позволяет получать несмещенные оценки вектора параметров θ по наблюдаемым входным и выходным данным. В [3, Теорема 2] построен критерий идентификации в форме ВФК для распределенной мультисенсорной системы фильтрации.

Пусть $D(\theta)$ – область определения параметра θ . Алгоритм параметрической идентификации строят путем реализации процедуры численной минимизации критерия ВФК, т. е. оценку $\hat{\theta}^*$ вычисляют по принципу

$$\hat{\theta}^* = \underset{\hat{\theta} \in D(\theta)}{\operatorname{argmin}} J_{\text{ВФК}}(\hat{\theta}). \quad (2)$$

Настоящая работа предлагает новый гибридный алгоритм идентификации параметра θ , суть которого в том, что оценка находится в два этапа:

1) вычисление начального приближения с помощью метаэвристического алгоритма оптимизации [5] и

2) уточнение найденной оценки с помощью алгоритма оптимизации градиентного типа [6]. Такой подход позволяет на первом этапе расширить пространство поиска, а на втором этапе – быстро уточнить найденную оценку.

Более того, для второго этапа мы впервые показываем, как вычислять градиент критерия ВФК по данным мультисенсоров, а именно, доказано следующее

Утверждение. Значения критерия ВФК и его градиента

$$\nabla_{\theta} J_{\text{ВФК}}(\hat{\theta}) = \left[\frac{\partial J_{\text{ВФК}}(\hat{\theta})}{\partial \theta_1} \quad \dots \quad \frac{\partial J_{\text{ВФК}}(\hat{\theta})}{\partial \theta_p} \right]^T$$

в алгоритме параметрической идентификации можно вычислить как

$$J_{\text{ВФК}}(\hat{\theta}) = \frac{1}{N} \sum_{k=1}^N [\varepsilon_k^-(\hat{\theta})]^2 \text{ и} \quad (3)$$

$$\frac{\partial J_{\text{ВФК}}(\hat{\theta})}{\partial \theta_j} = \left. \frac{\partial J_{\text{ВФК}}(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}} = \frac{2}{N} \sum_{k=1}^N \varepsilon_k^-(\hat{\theta}) \left. \frac{\partial \varepsilon_k^-(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}}$$

где $p=m+2$, $\hat{\theta}$ – значение параметра θ ,

$$\varepsilon_k^-(\hat{\theta}) = \frac{1}{m} \sum_{i=1}^m y_i(t_k) - s_k^-(\hat{\theta}) / \lambda_k^-(\hat{\theta}),$$

значения величин $s_k^-(\hat{\theta})$ и $\lambda_k^-(\hat{\theta})$ получены в распределенной схеме информационной калмановской фильтрации [3, разд. 4].

Доказательство основано на определении ВФК, правилах дифференцирования векторных функций и уравнениях распределенной калмановской фильтрации.

3. ЗАКЛЮЧЕНИЕ

В работе построен новый гибридный алгоритм параметрической идентификации распределенной мультисенсорной системы фильтрации, основанный на применении метода ВФК. Гибридность алгоритма означает наличие двух шагов:

1) вычисление начального приближения метаэвристическим алгоритмом и

2) уточнение найденной оценки градиентным алгоритмом.

Впервые показано, как вычислять градиент критерия ВФК в распределенной мультисенсорной системе. Алгоритм реализован и протестирован в среде MATLAB.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ и Правительства Ульяновской области в рамках научного проекта № 18-41-732002 р_мк.

ЛИТЕРАТУРА

- [1] Kordic, V. Kalman Filter / V. Kordic. – London, United Kingdom: IntechOpen, 2010. URL: <https://www.intechopen.com/books/4459>.
- [2] Mahmoud, M.S. Distributed Kalman filtering: a bibliographic review / M.S. Mahmoud, H.M. Khalid // IET Control Theory Appl. – 2013. – Vol. 7(4). – P. 483-501. DOI: 10.1049/iet-cta.2012.0732.
- [3] Семущин, И.В. Активная адаптация распределенной мультисенсорной системы фильтрации / И.В. Семущин, Ю.В. Цыганова // Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки. – 2019. – Т. 23, № 4. – С. 724-743. DOI: 10.14498/vsgtu1704.
- [4] Semushin, I.V. The APA based time-variant system identification / I.V. Semushin // Proc. of the 53rd IEEE Conference on Decision and Control. – 2014. – P. 4137-4141. DOI: 10.1109/CDC.2014.7040033.
- [5] Пантелеев, А.В. Метаэвристические алгоритмы глобальной оптимизации / А.В. Пантелеев, Д.В. Скавинская. – М.: Вузовская книга, 2019. – 332 с.
- [6] Fletcher, R. Practical Methods of Optimization / R. Fletcher. – Chichester, Great Britain: John Wiley & Sons Ltd, 1988. – 436 p.

Исследование погрешности разностного решения уравнения теплопроводности в многослойной среде методом вычислительного эксперимента

Ю.Ю. Кривошеева

Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
akinava.love@gmail.com

Аннотация—В данной работе рассмотрен метод нахождения погрешности разностного решения уравнения теплопроводности в трехслойной среде. Было показано, что существует возможность предсказания погрешности решения по размеру шагов сетки без численного расчета самого решения.

Ключевые слова— уравнение теплопроводности, разностная схема, погрешность решения.

1. ВВЕДЕНИЕ

Краевые задачи для уравнения теплопроводности – один из важных разделов уравнений математической физики. Помимо классических вариантов задач для однородной среды, для которых коэффициенты уравнения являются непрерывными, также интересны случаи, когда среда является неоднородной (включает в себя несколько слоев). К таким средам можно отнести многослойные обшивки технических конструкций, биологические объекты (кожа, сосуды).

Особый интерес представляют технические конструкции, подверженные тепловой нагрузке. Для таких конструкций нужно принимать во внимание вероятность влияния тепла на систему, так как тепловая нагрузка может привести к изменению ряда качеств материала, что приведет к выводу прибора из строя, его повреждению или же утрате весомых эксплуатационных качеств [1]. В связи с этим появляется необходимость разработки специальной тепловой защиты, которая позволила бы сохранить температуру внутри системы, обеспечить нормальное функционирование оборудования, прочность элементов конструкций.

Для уменьшения тепловой нагрузки, на поверхности, подверженные тепловому воздействию, наносят защитные покрытия. Таким образом, это сводится к задаче о теплопроводности слоистых структур, свойства теплопроводности которых меняются скачком.

2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ

Математическую модель задачи о распространении тепла в многослойном объекте можно записать в виде:

$$\begin{cases} c(x) \frac{\partial u(x, t)}{\partial t} = \frac{\partial}{\partial x} \left(k(x) \frac{\partial u(x, t)}{\partial x} \right), & 0 \leq x \leq L, 0 \leq t \leq T; \\ u|_{t=0} = u_c; \\ u|_{x=0} = u_c, \quad u|_{x=L} = \eta(t) + u_c, & 0 \leq t \leq T, \end{cases} \quad (1)$$

где $u(x, t)$ – непрерывная функция распределения температуры в объекте, $k(x)$ – кусочно-непрерывная функция коэффициента теплопроводности, $c(x)$ – кусочно-непрерывная функция удельной теплоемкости, L – длина объекта, T – время воздействия тепловой нагрузки.

Разностную неявную интегро-интерполяционную схему для задачи (1) запишем следующим образом:

$$\begin{cases} c_i \frac{u_i^{k+1} - u_i^k}{h_t} = \frac{1}{h_x} \left(a_{i+1} \frac{u_{i+1}^{k+1} - u_i^{k+1}}{h_x} - a_i \frac{u_i^{k+1} - u_{i-1}^{k+1}}{h_x} \right), \\ u_i^0 = u_c, \quad i = \overline{0, I}; \\ u_0^k = u_c, \quad u_i^k = \eta(t_k) + u_c, \quad k = \overline{1, K}. \end{cases} \quad (2)$$

Пусть 3 слоя соприкасаются в точках x_{01} и x_{02} и имеют характеристики $\tilde{c}, \tilde{k}; \hat{c}, \hat{k}$ и \check{c}, \check{k} соответственно. Тогда c_i и a_i в системе (2) будем вычислять следующим образом:

$$\begin{cases} c_i = \tilde{c}, & i < I_{01}; \\ c_i = \frac{\tilde{c} + \hat{c}}{2}, & i = I_{01}; \\ c_i = \hat{c}, & I_{01} < i < I_{02}; \\ c_i = \frac{\hat{c} + \check{c}}{2}, & i = I_{02}; \\ c_i = \check{c}, & i > I_{02}; \end{cases}$$

$$\begin{cases} a_i = \tilde{k}, & i \leq I_{01}; \\ a_i = \hat{k}, & I_{01} < i \leq I_{02}; \\ a_i = \check{k}, & i > I_{02}, \end{cases}$$

где $I_{01} = \frac{x_{01}}{h_x}, I_{02} = \frac{x_{02}}{h_x}$.

Для получения погрешности разложим разностное решение для шагов по пространству, отличающихся между собой в два раза:

$$u_{h_x, h_t} = [u]_{h_x, h_t} + Dh_t + Eh_x + O(h_t^2, h_x^2)$$

$$u_{\frac{h_x}{2}, \frac{h_t}{2}} = [u]_{\frac{h_x}{2}, \frac{h_t}{2}} + Dh_t + E \frac{h_x}{2} + O(h_t^2, \frac{h_x^2}{4})$$

где u – разностное решение, $[u]$ – точное аналитическое решение в узлах сетки, D, E – коэффициенты разложения.

Вычтем одно выражение из другого. При этом будем учитывать разность двух аналитических решений даст ноль. Таким образом получим часть погрешности $\Delta(h_x)$, связанную с измельчением шага по пространству:

$$\Delta(h_x) = \left| u_{h_x, h_t} - u_{\frac{h_x}{2}, \frac{h_t}{2}} \right| = E \frac{h_x}{2} + O(h_t^2, \frac{h_x^2}{4})$$

Для получения части погрешности, связанной с шагом по времени, проделаем аналогичные действия. В результате имеем формулу:

$$\Delta(h_t) = \left| u_{h_x, h_t} - u_{h_x, \frac{h_t}{2}} \right| = D \frac{h_t}{2} + O(\frac{h_t^2}{4}, h_x^2)$$

Используя коэффициенты E и D , можно предсказывать погрешность решения, не прибегая к расчетам самого решения. Это нужно для того, чтобы можно было, подбирая шаги дискретизации по пространству и времени, строить решение с интересующим уровнем погрешности.

3. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

В ходе численного расчета решения с параметрами $\tilde{c} = 1.15$ Дж/см³*К, $\hat{c} = 0.2$ Дж/см³*К, $\check{c} = 0.01$ Дж/см³*К, $\tilde{k} = 0.2$ Вт/см*К, $\hat{k} = 400$ Вт/см*К, $\check{k} = 0.09$ Вт/см*К, $l = 12$ см, $T = 20$ с, $\eta(t) = 1000 \sin \frac{\pi t}{2T}$, $u_c = 300$ К и различными шагами по пространству и времени и последующим расчетом долей погрешностей были получены коэффициенты разложения E и D , что позволило установить формулу для теоретического расчета полной погрешности при данных параметрах:

$$\Delta = 16.79h_x + 12.75h_t \quad (3)$$

В таблице 1 представлено отклонение ε полной теоретической погрешности $\Delta_{\text{теор}}$, рассчитанной по формуле (3) от экспериментальной $\Delta_{\text{эксп}}$, полученной при численном моделировании.

ТАБЛИЦА 1. ОТКЛОНЕНИЕ СУММАРНОЙ ПОГРЕШНОСТИ ДЛЯ РЕШЕНИЯ В ТРЕХСЛОЙНОЙ ОБЛАСТИ

h_t	h_x	$\Delta_{\text{теор}}$	$\Delta_{\text{эксп}}$	ε
0.08	0.8	14.59	14.64	0.34%
0.06	0.6	10.95	10.72	1.23%
0.05	0.5	9.12	9.06	0.65%
0.04	0.4	7.29	7.36	0.96%
0.03	0.3	5.47	5.38	1.65%
0.02	0.2	3.65	3.60	1.37%
0.01	0.1	1.82	1.86	2.19%

Из таблицы видно, что отклонение суммарной теоретической погрешности от экспериментальной остается малым, что позволяет делать весьма точный прогноз. Таким образом, мы можем получать решение с заданным уровнем погрешности, задавая шаги.

4. ЗАКЛЮЧЕНИЕ

Таким образом, в ходе численного эксперимента была получена погрешность разностного решения задачи теплопроводности в многослойной области при различных шагах сетки. На основе полученных результатов была построена формула для предсказания погрешности по заданным шагам и показано, что по этой формуле возможно получать интересующий уровень погрешности с достаточно высокой точностью.

ЛИТЕРАТУРА

- [1] Танана, В.П. О решении обратной граничной задачи для композитных материалов / В.П. Танана, А.А. Ершова // Вестник удмуртского университета. – 2018. – Т. 28, № 4. – С. 474–488. DOI: 10.20537/vm180404.
- [2] Шашков, А.Г. Тепло- и массообмен в потоке нагретого газа / А.Г. Шашков. – Минск: Изд-во «Наука и техника», 1974. – 101 с.
- [3] Фихтенгольд, Г.М. Курс дифференциального и интегрального исчисления: в 3 т. / Г.М. Фихтенгольд. – М.: Физматлит, 2001. – Т. 3. – 662 с.
- [4] Тихонов, А.Н. Уравнения математической физики: учебное пособие для университетов / А.Н. Тихонов, А.А. Самарский. – М.: Наука, 1977. – 735 с.
- [5] Сабитов, К.Б. Уравнения математической физики: учеб. пособие / К.Б. Сабитов. – М.: Физматлит, 2013. – 352 с.
- [6] Самарский, А.А. Введение в численные методы: учеб. пособие для вузов / А.А. Самарский. – М.: Наука, 1987. – 288 с.
- [7] Дегтерев, А.А. Метод конечных разностей: учеб. пособие / А.А. Дегтерев. – Самара: Самарский университет, 2019. – 81 с.

Agent-Oriented simulation modeling of systems

M.A. Nikitina
V.M. Gorbatov Federal Research Centre for Food Systems of RAS
Moscow, Russia
nikitinama@yandex.ru

Abstract—A situational model of identification of a state of a complex technological system under conditions of uncertainty is proposed based on a structural and parametric description of the system and agent technologies. The methodology of developing a self-learning intelligent agent capable of identifying the current situation with incomplete and fuzzy information and making adequate decisions on its normalization in real time in the management of a processing plant's technological system is described.

Keywords— *situational analysis, multi-agent modeling, technological system.*

1. INTRODUCTION

An intelligent agent [1] is a simulation model of an active element capable of performing certain assigned functions of a living or cybernetic organism, depending on the behavior of other agents and environmental influences.

Self-learning purposeful agents are able to accumulate knowledge based on current data and the ontology of events in the process of interaction with other agents and the environment, adapt to the situation, choose a strategy for achieving the chosen goal and assess the degree of its achievement.

The general algorithm of behavior of an intelligent agent (see "Fig. 1") includes the following steps: identifying the situation, assessing its own state, and correcting the goal, followed by a reflexive reaction or meaningful (intellectual) decision-making towards the goal.

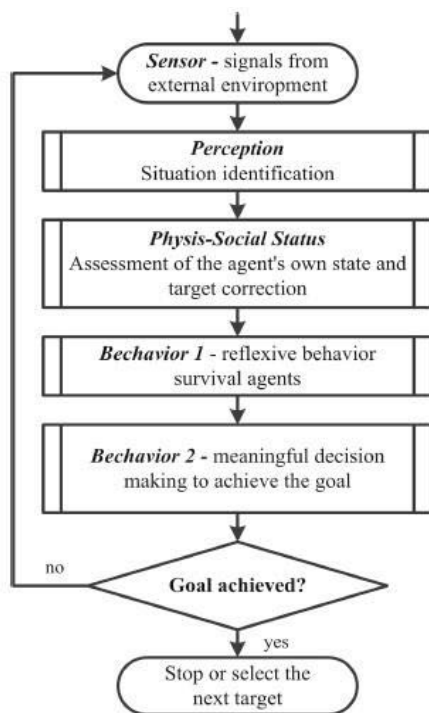


Fig. 1. General algorithm for the behavior of an intelligent agent

The criterion of agent's intelligence is the degree of completeness and depth of a priori knowledge, learning strategies, and decision-making algorithms in conditions of uncertainty, risk, and conflict.

Each operation has its own algorithmic and software module that provides:

- Perception of information and accumulation of knowledge about the external environment and the environment of interaction or conflict (sensory module).
- A mechanism of interaction and processing of data from counterparties.
- Analysis of own and counterparty status with selection or correction of objective functions (intelligent module).
- Making autonomous decisions and choosing strategies.

The behavior of the agent can be represented by some recursive form which describes the finding and choosing at the next step of a function the transition from the initial state to a new state towards an improved objective function.

In some cases, such a problem can be solved by mathematical programming methods with correction of the objective function and individual constraints at the next step of changing the agent's state depending on the situation and the distance to the goal in conditions of uncertainty and fuzzy information.

The parametric description of an agent includes a set of goals and a knowledge base in a particular area, a vector of characteristics of its state, a bank of models and behavior strategies, and a description of external relations with agents and the environment.

2. THE METHODOLOGY OF AGENT-BASED MODELING

The methodology of agent-based modeling of a self-learning agent is reduced to the following stages.

- Parametric description of the external environment of agent's activity with formalization of a set of factors influencing the functional state and objective function of the agent in situational decision-making conditions.
- Parametric description of functional blocks of a technological system in the form of a set of vectors of input and output factors, state parameters, and an objective function.
- Description of an autonomous intelligent agent with multiple state variables, input and sensor variables that communicate with other agents and the environment.
- Development of a mathematical model of agent training and an algorithm for its behavior with training procedures and identification of current situations and

decision-making in the form of a discrete-event description and decision-making strategies in conditions of sufficient, incomplete and fuzzy information.

- Software implementation of the agent model in an agent-oriented model description language, in the form of a basic component [2, 6] of a multi-agent model of a technological system.

The basic component includes the sections describing the agent's state variables DECLARATION OF ELEMENTS and the dynamics of changes in its state (behavior) DYNAMIC BEHAVIOR with an analytical or discrete-event description. The syntactic form of the basic component is as follows (see "Fig. 2").

```
BASIC COMPONENT < name >
[mobile_subclass_declaration - mobile components]
[subunit_declaration - base units]
[local_definitions - arrays, functions, distribution laws]

DECLARATION OF ELEMENTS
[list_of_constants - constants]
[list_of_state_variables - state variables]
[list_of_dependent_variables - calculated variables]
[list_of_sensor_variables - sensory variables]
[list_of_random_variables - random variables]
[list_of_transition_indicators - transition indicators]
[list_of_sensor_indicators - touch indicators]
[list_of_locations - storage arrays]
[list_of_sensor_locations - sensor drives]

DYNAMIC BEHAVIOUR
| algebraic_equation - algebraic equations
| differential_equations - differential equations
| region_defining_statement - areas of certain states
| event_defining_statement - events

END
END OF < name >
```

Fig. 2. The syntactic form of the basic component

- Practical implementation of agent-based technologies in a universal simulation system that provides an experimentation environment, an agent-oriented model description language, and software tools for organizing experiments [2, 6].

- An easy-to-learn non-commercial tool for implementing agent-based technologies and multi-agent simulation modeling is the Simplex3 universal simulation system with its own object-oriented language Simplex3-MDL (Modell Deskription Language) [1, 6] for describing system-dynamic, discrete-event, and multi-agent models. The Simplex3 system, developed at the Universities of Nuremberg-Erlangen, Passau and Magdeburg (Germany), is intended for teaching the art of modeling and for non-commercial use in research and development.

- The main advantage of the Simplex3-MDL language is the simplicity of description of agents and multi-agent models of large systems of any physical, social, or biological nature in a form close to natural parametric descriptions of the state of agents, their sensory connections with other agents, and the dynamics of state changes using well-known mathematical tools, and discrete-event descriptions.

In our country, the most widespread modeling system is the universal modeling system Anylogic [3], developed by the Russian company XJ Technologies and based on modern information technologies with graphical model construction and programming in the object-oriented language Java.

3. CONCLUSION

Agent technologies with neural network algorithms for the behavior of self-learning intelligent agents with recognition of current situations in conditions of fuzzy information, uncertainty and risk open up a new direction for the intellectualization of expert computer decision support systems in complex technological process management systems, as well as in virtual research of the influence of various technological factors on abnormal system states.

For software implementation of a self-learning intelligent agent, the *Simplex3* universal simulation system can be used with the specialized object-oriented language *Simplex3-MDL (Modell Deskription Language)* for describing system-dynamic, discrete-event, and multi-agent models [1]. The procedure for training an agent in the dynamics of its behavior is based on a multi-layer neural network with pairs of interconnected input and output vectors and recurrent tuning of synaptic connections based on similarity (for example, Hamming distance). The training sample vectors are formed using a structural-parametric situational model and an algorithm for causal identification of a technological system in conditions of certainty [4, 5].

The proposed direction of intellectualization of the situational modeling of systems lays the foundation for the construction of intelligent expert systems (IES) for making optimal decisions and operational management of food quality and safety at various stages of production at processing enterprises of the agro-industrial complex [4, 5].

ACKNOWLEDGMENT

This article is published as part of scientific research theme No. 0585-2019-0008 under the state assignment of the federal state budgetary scientific institution 'V. M. Gorbatov's Federal Research Centre for Food Systems' of RAS.

REFERENCES

- [1] Ivashkin, Yu.A. Multiagent modeling in Simplex3 simulation system: Textbook / Yu.A. Ivashkin. – Moscow: "Laboratory of Knowledge" Publisher, 2016. – 350 p.
- [2] Schmidt, B. The Art of Modelling and Simulation / B. Schmidt. – Chent: "SCS-Europe BVBA" Publisher, 2001. – 480 p.
- [3] Karpov, Yu.G. Simulation of systems. Introduction to modeling with AnyLogic 5 / Yu.G. Karpov. – Saint Petersburg: "BHV-Saint Petersburg" Publisher, 2005. – 400 p.
- [4] Ivashkin, Y.A. Neural network and agent technologies in the structural-parametric modeling of technological systems / Y.A. Ivashkin, I.G. Blagoveschensky, M.A. Nikitina // CEUR Workshop Proceedings. Proceedings of the School-Seminar on Optimization Problems and their Applications. – 2018. – P. 169-180.
- [5] Nikitina, M. Agent-Based Situational Modeling and Identification Technological Systems in Conditions of Uncertainty / M. Nikitina, Y. Ivashkin // Advances in Intelligent Systems and Computing. – 2020. – Vol. 1156 AISC. – P. 109-119.
- [6] Simplex 3 [Electronic resource]. – Access mode: <http://www.simplex3.net> (01.02.2022).

Разработка нагрузочного теста для исследования производительности распределенных файловых систем на основе подсчета ортогональных пар диагональных латинских квадратов

С.Н. Попов

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
stef4n.popov@gmail.com

С.В. Востокин

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
eastst@mail.ru

Аннотация—В статье описан опыт разработки нагрузочного теста для исследования производительности распределенных файловых систем. В качестве предмета исследования использовался подсчет ортогональных пар диагональных латинских квадратов. Выполнена программная реализация данного теста, пригодная для исполнения на произвольных невыделенных вычислительных ресурсах сети Интернет. Работоспособность программной реализации теста проверена в конфигурации на VPN-сервере с установленной системой TLJH, пользователи которой имитировали вычислительные ресурсы. Экспериментально проверена корректность выполнения тестов для различного количества разбиений массива ДЛК на файлы, а также приведен результат ускорения.

Ключевые слова— *распределенная файловая система, диагональные латинские квадраты, невыделенные вычислительные ресурсы, TLJH, VPN-сервер.*

1. ВВЕДЕНИЕ

Одним из направлений сокращения стоимости вычислений является использование так называемых невыделенных (non-dedicated) вычислительных ресурсов. Примерами таких ресурсов являются: машины добровольцев, предоставляемые ими временно для научных проектов [1]; свободные от основной нагрузки узлы кластеров [2] или суперкомпьютеров [3]; бесплатные или недорогие виртуальные машины облачных провайдеров [4]. Для учета специфики перечисленных окружений в вычислительных приложениях (compute-intensive) достаточно разбить вычисления на множество задач. Однако для использования невыделенных ресурсов в задачах обработки данных (data-intensive) необходимо также подобрать систему распределенного хранения, так как производительность системы хранения будет оказывать существенное влияние на эффективность вычислений в целом.

В связи с этим актуальными являются вопросы о выборе подходящей системы распределенного хранения, предсказания производительности приложения с выбранной системой, тонкой настройки приложения и т. п.

В качестве подхода к решению перечисленных вопросов в работе рассматривается нагрузочный тест и

его программная реализация, построенные на основе задачи подсчета пар диагональных латинских квадратов 7-ого порядка [5].

2. МЕТОДИКА ПОСТРОЕНИЯ ТЕСТА

Идея теста состоит в попарном сопоставлении специальных математических объектов — нормализованных диагональных латинских квадратов (ДЛК) 7-ого порядка. Пару ДЛК можно объединить в новый квадрат, каждый элемент которого получен объединением элементов исходных ДЛК в совпадающих позициях. При этом новый квадрат может также обладать свойством ДЛК: не содержать повторяющихся элементов во всех строках, столбцах, главной и побочной диагоналях. В этом случае говорят об ортогональности исходной пары ДЛК.

Для построения теста существенно, что вычислительный процесс, реализующий подсчет ортогональных пар латинских квадратов, (а) моделирует часто встречающийся на практике типовой информационный процесс попарной обработки данных в независимых задачах AllPairs; (б) имеет простой критерий корректности результата, так как число пар для ДЛК данного порядка известно; (в) существенно зависит от производительности от внутреннего алгоритма работы системы хранения; (г) массив данных для теста, включающий 171200 нормализованных ДЛК 7-ого порядка, уникален и генерируется программно.

Тривиальная стратегия хранения данных — размещение файлов с ДЛК на файловом сервере на некотором узле. При этом вычислительные узлы, выполняющие сопоставление пар ДЛК, будут каждый раз запрашивать файлы с файлового сервера. Простая оптимизация заключается в кэшировании данных на вычислительных узлах после их получения с файлового сервера, что позволяет избежать повторных передач файлов, если они переданы на узел ранее для другой задачи. Еще одной оптимизацией является прямая передача данных между кэшами систем хранения вычислительных узлов, минуя узел файлового сервера, что частично разгружает его канал передачи данных. Производительность процесса AllPairs улучшится за счет перечисленных оптимизаций, если они автоматически поддерживаются исследуемой системой хранения данных.

Тест проводится для различных разбиений множества ДЛК на файлы: все ДЛК в одном файле, ДЛК в 2х файлах, ДЛК в 3-х файлах и так далее. Файлы с ДЛК содержат примерно равное (разница не более чем на один квадрат) число квадратов. В тестах измеряется среднее время выполнения поиска ортогональных пар ДЛК. Тестирование разных разбиений множества ДЛК проводится до тех пор, пока не будет достигнут минимум времени вычисления. Снижение времени должно наблюдаться по причине распараллеливания, а повышение времени вычисления после достижения минимального значения происходит из-за снижения трудоёмкости обработки, приходящейся на единицу обрабатываемых данных (гранулярность) и, следовательно, росту доли накладных затрат на единицу обрабатываемых данных. Лучшая система распределенного хранения обеспечивает минимум времени счета и большее ускорение при прочих равных условиях, а также эффективнее работает с менее гранулярными задачами.

Разработанное приложение для тестирования использует генератор ДЛК, предоставленный авторами работы; подсистему запуска задач на основе Templet SDK [6, 7]; сервер платформы Everest [8, 9], распределяющий задачи среди программ-агентов Everest, установленных на ресурсах. Также на ресурсах устанавливаются клиенты исследуемой системы хранения. Развертывание и тестирование описанной программной инфраструктуры проведено на VPN-сервере с установленной системой TLJH [10]. Для имитации ресурсов использовались отдельные сессии TLJH, для имитации системы хранения использовался разделяемый пользователями TLJH каталог файловой системы VPN-сервера. Экспериментально проверена корректность выполнения тестов для различного количества разбиений массива ДЛК на файлы; получено ускорение 1.5 для разбиения массива на 2 части на 10 виртуальных процессорах.

3. ЗАКЛЮЧЕНИЕ

В работе предложен нагрузочный тест для исследования производительности распределенных файловых систем на основе алгоритма поиска ортогональных пар диагональных латинских квадратов 7-ого порядка. Выполнена программная реализация данного теста, пригодная для исполнения на

произвольных невыделенных вычислительных ресурсах сети Интернет. Работоспособность программной реализации теста проверена в конфигурации на VPN-сервере с установленной системой TLJH, пользователи которой имитировали вычислительные ресурсы. В дальнейшем тест планируется использовать для сравнительного исследования системы хранения Everest и системы IPFS [11] на предмет применимости в практических задачах обработки данных на невыделенных вычислительных ресурсах.

ЛИТЕРАТУРА

- [1] Mengistu, T. Survey and Taxonomy of Volunteer Computing / T. Mengistu, D. Che // *ACM Computing Surveys*. – 2019. – P. 52. DOI: 10.1145/3320073.
- [2] Reddy, C.K.K. High Performance Computing Cluster System and its Future Aspects in Processing Big Data / C.K.K. Reddy, K.E. Bala Chandrudu, P.R. Anisha, G.V.S. Raju // *International Conference on Computational Intelligence and Communication Networks (CICN)*. – 2015. – P. 881-885. DOI: 10.1109/CICN.2015.173.
- [3] Ungurean, I. High-performance computing on a supercomputer based on new-generation processors / I. Ungurean, I. Rusu, S. Pentiu / *5th Romania Tier 2 Federation Grid, Cloud & High Performance Computing Science (RQLCG)*. – 2012. – P. 96-99.
- [4] Roloff, E. High Performance Computing in the cloud: Deployment, performance and cost efficiency / E. Roloff, M. Diener, A. Carissimi, P.O.A. Navaux // *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings*. – 2012. – P. 371-378. DOI: 10.1109/CloudCom.2012.6427549.
- [5] Kochemazov, S.E. Fast Algorithm for Enumerating Diagonal Latin Squares of Small Order / S.E. Kochemazov, E.I. Vatutin, O.S. Zaikin / *ArXiv preprint: 1709.02599*, 2017.
- [6] Vostokin, S.V. Templet: A markup language for concurrent actor-oriented programming / S.V. Vostokin // *CEUR Workshop Proceedings*. – 2016. – Vol. 1638. – P. 460-468.
- [7] The Templet Project [Electronic resource]. — Access mode: <https://github.com/the-templet-project> (06.02.2022).
- [8] The Everest Project [Electronic resource]. — Access mode: <http://everest.distcomp.org> (06.02.2022).
- [9] Sukhoroslov, O. Web-Based Platform for Publication and Distributed Execution of Computing Applications / O. Sukhoroslov, S. Volkov, A.A. Afanasiev // *14th International Symposium on Parallel and Distributed Computing (ISPDC)*. IEEE. – 2015. – P. 175-184.
- [10] The Littlest Jupyter rHub project [Electronic resource]. — Access mode: <https://tljh.jupyter.org> (06.02.2022).
- [11] Muralidharan, S. An InterPlanetary File System (IPFS) based IoT framework / S. Muralidharan, H. Ko. // *IEEE International Conference on Consumer Electronics (ICCE)*. – 2019. – P. 1-2. DOI: 10.1109/ICCE.2019.8662002.

Выявление признаков объекта на основе неполных и противоречивых эмпирических данных

В.А. Семенова

Самарский федеральный исследовательский центр РАН
Институт проблем управления сложными системами РАН
Самара, Россия
queenbfjr@gmail.com

С.В. Смирнов

Самарский федеральный исследовательский центр РАН
Институт проблем управления сложными системами РАН
Самара, Россия
smirnov@iccs.ru

Аннотация—В работе рассматривается проблема построения фундаментальной семантической модели исследуемого объекта, которая определяет его признаки, т.е. состав присущих объекту свойств. Набор этих признаков является подмножеством всех измерившихся (в широком смысле) у исследуемого объекта свойств. Обработка результатов подобных многомерных измерений, наблюдений, регистрации экспертных мнений и т.п. дает исходный эмпирический материал в форме в общем случае логически многозначных оценок истинности суждений вида «объекту x присуще свойство y ». С другой стороны, субъект может располагать априорным знанием о вариантах сосуществования различных признаков у исследуемого объекта. Для таких обстоятельств предлагается методика выявления всех допустимых наборов признаков объекта, включающая логико-комбинаторный метод выбора «лучшего» из этих наборов, который и принимается в качестве семантической модели объекта. В работе рассматриваются примеры приложений разработанных моделей и методов, указываются задачи дополнительных исследований для обоснования предлагаемой методики семантической идентификации объекта на основе эмпирических данных.

Ключевые слова— *признаковое описание объекта, достоверность эмпирических данных, ограничения существования, нормальное множество.*

1. ВВЕДЕНИЕ

Знания о предметной области (ПрО) в классической для современной науки и техники субъект-объектной оппозиции приобретаются путем анализа данных многомерных измерений, наблюдений, регистрации экспертных мнений и т.п. При этом воспроизводится традиционный процесс познания в форме выдвижения и последующей проверки гипотез. Субъект формирует состав измерительных (в широком смысле) процедур, фиксируя тем самым гипотезы о признаковом описании объекта, а результаты измерений в общем случае часть этих гипотез подтверждают, а другую их часть опровергают. Последнее происходит в силу того, что полное определение всякой процедуры измерения предполагает включение в её всегда ограниченный диапазон измерений константы «несколько», «ничто» [1], получение которой в качестве результата измерения свидетельствует об ошибочности включения измерившегося свойства объекта в его признаковое описание. Аналогично проявляет себя концептуальное шкалирование [2] полученных результатов измерений.

В работе предлагается методика выявления признаков исследуемого объекта ПрО, учитывающая реалии формирования исходных объектно-признаковых данных.

2. МЕТОД ИДЕНТИФИКАЦИИ ПРИЗНАКОВОГО ОПИСАНИЯ ОБЪЕКТА

На практике обнаруживаются два обстоятельства, которые усложняют процесс идентификации признакового описания объекта:

- при выдвижении гипотез о возможном признаковом описании объекта на множестве измеряемых свойств могут возникать так называемые ограничения существования – бинарные отношения несовместимости, обусловленности (последнее может приводить к взаимообусловленности) признаков [3, 4] (Рис. 1);
- накопление эмпирической информации в результате выполнения в общем случае многих серий измерений свойств у исследуемого объекта, использования конкурирующих процедур измерения одного и того же свойства, учета достоверности серий измерений и степеней доверия к процедурам и др. приводит к неполным и противоречивым данным о присущих объекту свойствах. Интерпретация и консолидация таких данных для оценки истинности суждений вида «объекту x присуще свойство y » требует привлечения моделей и методов многозначной логики [5].

Идея предлагаемого метода состоит в том, что при наличии обоих факторов задачу получения признакового описания исследуемого объекта, удовлетворяющего ограничениям существования, можно решить путем отказа от части эмпирической информации, имеющей при прочих равных условиях меньшее значение их агрегированного показателя достоверности. Такая нормализация признакового описания, во-первых, требует исключить из него члены, обуславливающие признаки, согласно измерениям отсутствующие в описании объекта, а, во-вторых, если в описании имеются подмножества несовместимых признаков, сохранить в каждом таком подмножестве не более одного признака. Именно в последнем случае возникает задача минимизации агрегированного показателя достоверности исключаемых эмпирических свидетельств.

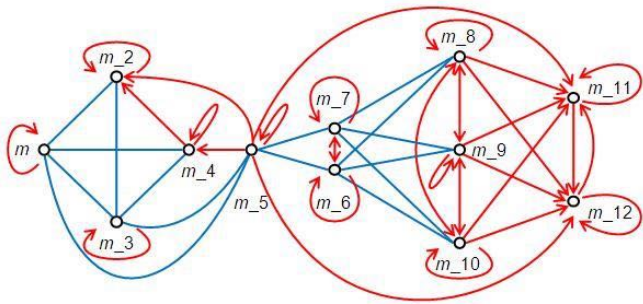


Рис. 1. Пример описания ограничений существования в форме отношений обусловленности и несовместимости на множестве измеряемых свойств: вершины графа – свойства, дуги – обусловленность (двунаправленная дуга - обозначение 2-х встречных дуг), ребра – несовместимость свойств

Анализ свойств симметрии бинарных отношений, определяющих ограничения существования, позволяет при выявлении нормального множества признаков объекта заменить рассмотрение статуса отдельных измеряемых свойств анализом статуса их характерных сопряженных групп [6] (Рис. 2, 3), что в общем случае существенно повышает производительность предлагаемого метода.

3. ПРИЛОЖЕНИЯ РАЗРАБОТАННОЙ МЕТОДИКИ

Выявление признаков объекта с учетом ограничения существования свойств на основе неполных и противоречивых данных имеет широкую сферу приложения. В работе обсуждается применение предлагаемого подхода в задачах онтологического анализа данных, формирования когнитивных карт, портретной идентификации людей в криминалистике, структурного синтеза проектно-конструкторских решений.

4. ЗАКЛЮЧЕНИЕ

Разработанная методика идентификации признаков описания исследуемого объекта опирается на совокупность хорошо обоснованных математических моделей и строго доказанных утверждений. Включенная в неё задача логико-комбинаторной оптимизации допускает эффективное решение рекурсивным алгоритмом сокращенного обхода конечного бинарного дерева.

К числу перспективных дополнительных задач обоснования предлагаемой методики можно отнести:

- исследование влияния различных способов агрегированной оценки достоверности денонсируемых эмпирических данных на характеристики получаемого признакового описания объекта;
- количественное оценивание сокращения вычислительных затрат, достигаемого благодаря трансформации описания ограничений существования;
- сравнение предлагаемой методики с ранее разработанным подходом к идентификации признакового описания объекта на основе эвристики [7].

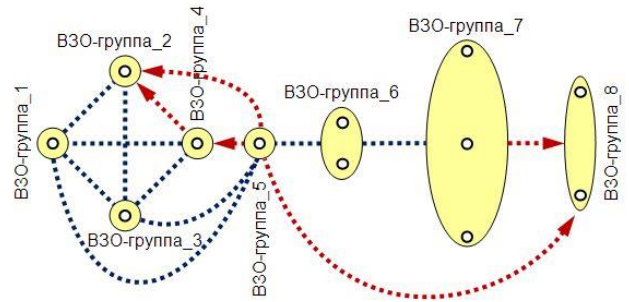


Рис. 2. Пример описания ограничений существования в форме обобщенных отношений обусловленности и несовместимости на множестве групп взаимобусловленных измеряемых свойств (ВЗО-групп) – описание обозначений см. на Рис. 1

	m	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_10	m_11	m_12
ВЗО-группа_1	o											
ВЗО-группа_2		o										
ВЗО-группа_3			o									
ВЗО-группа_4				o								
ВЗО-группа_5					o							
ВЗО-группа_6						o	o					
ВЗО-группа_7								o	o	o		
ВЗО-группа_8											o	o
О-группа_1		o		x								
О-группа_2		o			x							
О-группа_3				o	x							
О-группа_4					x							
О-группа_5								x	x	x	o	o
Н-группа_1	x		x		x							
Н-группа_2	x		x	x								
Н-группа_3	x	x										
Н-группа_4					x	x_6	x_6					
Н-группа_5						x_6	x_6	x_7	x_7	x_7		

Рис. 3. Пример соответствия «Группы сопряженных свойств – свойства». В группе сопряженных свойств вида «О-группа» одна группа взаимобусловленных свойств (ВЗО-группа) обуславливает другую ВЗО-группу. Группа сопряженных свойств вида «Н-группа» включает в себя по меньшей мере две ВЗО-группы; знаки инцидентности, помеченные одинаковым номером следует считать одной инцидентностью с ВЗО-группой, идентифицированной тем же номером

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Минобрнауки России, коды научных тем FMRW-2019-0023 и FMRW-2022-0030.

ЛИТЕРАТУРА

- [1] Смирнов, С.В. Онтологический анализ предметных областей моделирования / С.В. Смирнов // Известия СамНЦ РАН. – 2001. – Т. 3, № 1. – С. 62-70.
- [2] Ganter, B. Conceptual scaling / B. Ganter, R. Wille // Applications of Combinatorics and Graph Theory to the Biological and Social Sciences. – N.Y.: Springer-Verlag, 1989. – P. 139-167. DOI: 10.1007/978-1-4684-6381-1_6.
- [3] Lammari, N. Building and maintaining ontologies: a set of algorithms / N. Lammari, E. Metais // Data & Knowledge Engineering. – 2004. – Vol. 48(2). – P. 155-176. DOI: 10.1016/S0169-023X(03)00103-4.
- [4] Пронина, В.А. Использование отношений между атрибутами для построения онтологии предметной области / В.А. Пронина, Л.Б. Шипилина // Проблемы управления. – 2009. – № 1. – С. 27-32.
- [5] Smirnov, S.V. Biconstituent phenomenon of information and cognitive data analysis / S.V. Smirnov // Procedia Engineering. – 2017. – Vol. 201. – P. 773-778. DOI: 10.1016/J.PROENG.2017.09.626.
- [6] Семенова, В.А. Алгоритмизация формирования и прагматической трансформации ограничений существования свойств предметной области / В.А. Семенова, С.В. Смирнов // Онтология проектирования. – 2020. – Т. 10, № 3(37). – С. 361-379. DOI: 10.18287/2223-9537-2020-10-3-361-379.
- [7] Samoilov, D.E. Defuzzification of the initial context in Formal Concept Analysis / D.E. Samoilov, V.A. Semenova, S.V. Smirnov // CEUR Workshop Proceedings. – 2019. – Vol. 2416. – P. 1-9. DOI: 10.18287/1613-0073-2019-2416-1-9.

Разработка информационной системы для повышения эффективности эксплуатации нефтяных скважин

Г.В. Спиридонов
Казанский национальный
исследовательский технический
университет им. А. Н. Туполева -
КАИ
Казань, Россия
grisha5991@gmail.com

В.В. Мокшин
Казанский национальный
исследовательский технический
университет им. А. Н. Туполева -
КАИ
Казань, Россия
vladimir.mokshin@mail.ru

Э.Х. Шамсиев
Альметьевский филиал Казанского
национального исследовательского
технического университета им. А.
Н. Туполева - КАИ
Казань, Россия
EKhShamsiev@stud.kai.ru

Аннотация—В данный момент грамотное и эффективное управление нефтегазовым производством является одной из наиболее важных и, в то же время, непростых задач. В ходе работы сотрудники производства сталкиваются с огромным количеством проблем, среди которых особенно злободневными являются низкий уровень планирования и нерелевантное прогнозирование процессов. Целью данной работы является разработка АИС для повышения эффективности эксплуатации нефтяных скважин за счёт улучшения качества планирования производственно-технических мероприятий.

Ключевые слова— прогнозирование, планирование, анализа состояния нефтяных скважин, информационная система.

Планирование можно определить как функцию, отвечающую за постановку целей и задач производства, пути их достижения и выявление необходимых для этого ресурсов. Проблема неэффективного планирования и прогнозирования возникла отчасти из-за устаревания существующих технических систем, неиспользования в полной мере объёма информации и недостаточного общего уровня квалификации персонала предприятий. Выявленные факторы способствовали необходимости создания автоматизированной информационной системы (далее – АИС).

В системе мероприятий план рассматривается как вариант наиболее оптимального использования ресурсов предприятия, обеспечивающий достижение целей и выполнение задач производства, а прогнозирование – как предвидение развития производства и нахождение решений, которые обеспечивают наилучший вариант перспективы дальнейшей деятельности предприятия. Поэтому эксплуатационные мероприятия – это, прежде всего, мероприятия, осуществляемые в процессе использования оборудования нефтегазодобычи, направленные на поддержание их технического состояния на уровне заданных экологических и производственных нормативов.

В работе получили отражение как теоретические, так и практические решения ресурсосбережения на основе сформулированных задач, что является исключительно важным аспектом для управления. Логика разработанной информационной системы выстроена от частного к общему.

В работе решаются следующие задачи:
- анализ состояния предметной области;

- определение характеристик скважин;
- формирование эксплуатационных мероприятий для нефтяных скважин;
- разработка программного интерфейса для управления АИС.

Предлагаемая автоматизированная информационная система может не только повысить качество принимаемых решений, но и увеличить прибыль производства и объёмы добываемой нефти. В представленной работе определяется комплекс мероприятий, необходимых для дальнейшей эксплуатации нефтяных скважин.

Анализ данных осуществляется с помощью статистических методов исследования. В ходе обработки определяются зависимости между показателями эффективности и производственно-техническими факторами, что является востребованным и новым в современных условиях нефтегазовых разработок.

Планирование является одним из первых и наиболее серьёзных этапов управления объектом. В процесс планирования вовлекаются сотрудники различных подразделений, а потому описание технологии принятия решений на каждом уровне задач удобно представлять в виде диаграмм.

Для достижения поставленной цели необходимо решить следующие задачи:

1. определить характеристики скважин. Эта задача может быть решена с помощью математических методов и компьютерного обеспечения, сокращающих необходимое время для анализа зависимостей между результативными показателями и производственно-техническими факторами и повышающих достоверность полученных результатов.

2. сформировать эксплуатационные мероприятия нефтяных скважин.

Для наглядного представления моделирования информационного обмена между системой и внешним миром и/или между частями этой системы могут быть использованы диаграммы потоков данных Data Flow Diagrams (DFD).

3. разработать программный интерфейс. Результатом внедрения АИС и выполнения поставленных задач являются данные, на основании которых могут быть приняты реальные управленческие решения. Главными потребителями таких решений окажутся инженерные структуры нефтедобывающих компаний.

Построение математической модели состоит из следующих этапов:

1. Сбор исходных статистических данных (далее – ИСД).
2. Проверка ИСД на достоверность.
3. Временное прогнозирование.
4. Корреляционный анализ и оценка коэффициентов корреляции.
5. Проведение множественного регрессионного анализа.
6. Формирование мероприятий и обучение нейронной сети.

Для прогнозирования временных рядов воспользуемся методом авторегрессии и проинтегрированного скользящего среднего (далее – АРПСС).

Для прогнозирования данных воспользуемся регрессионным анализом. Данный статистический метод исследования влияния регрессоров (независимых переменных) на зависимые переменные основывается на методе наименьших квадратов (далее – МНК). МНК – это метод, требующий минимизации суммы квадратов отклонений искомых функций от имеющихся переменных.

Для разработки автоматизированной информационной системы было использовано следующее прикладное программное обеспечение:

1. СУБД SQL Server 2014;
2. Microsoft Visual Studio 2015;
3. Пакет прикладных программ IBM SPSS Statistics 26;
4. Табличный редактор Microsoft Excel 2016;
5. CASE-средства: ERwin, BPwin.

Все описанные математические алгоритмы были внедрены в соответствующий программный продукт, а все расчёты происходили автоматически на основе данных.

В ходе разработки автоматизированной информационной системы были выполнены следующие задачи по поддержке актуального состояния базы данных:

- 1) ведение базы данных переменных;
- 2) ведение базы данных по ИСД;
- 3) ведение базы данных по ОСХ;
- 4) ведение базы данных по данным проверки на нормальное распределение;
- 5) ведение базы данных по данным результата корреляционного анализа;
- 6) ведение базы данных по данным результата временного прогноза;
- 7) ведение базы данных по данным регрессионного уравнения;
- 8) ведение базы данных по данным обучения нейронной сети и формированию мероприятий.

На первом этапе были определены сущности и указаны соответствующие им атрибуты, домены и типы данных. Далее производилось построение логической модели базы данных по исходной концептуальной модели.

Таким образом, в процессе разработки АИС была проанализирована серия математических и информационных методов, позволяющих осуществить

прогноз производственно-технических показателей нефтяных скважин и произвести планирование мероприятий для их эффективного использования. На основе полученных данных были найдены и теоретически подтверждены возможные решения поставленных задач.

Конечным продуктом работы является апробация автоматизированной информационной системы, позволяющей объединить все этапы планирования и прогнозирования. Всё это представляет готовую схему эксплуатационных мероприятий с учётом имеющихся в настоящий момент потребностей у нефтегазодобывающих производств. Благодаря использованию программного обеспечения становится возможным применять обучение нейронных сетей при решении задач планирования мероприятий, что является чрезвычайно актуальным и перспективным направлением в сфере информационных технологий.

При внедрении данной информационной системы нефтегазовые производства смогут рассчитывать на повышение эффективности использования нефтяных скважин, оптимальное проведение мероприятий по эксплуатации производственных комплексов и, как следствие этого, на увеличение объёмов добычи нефтегазовых ресурсов и получение дополнительной прибыли.

ЛИТЕРАТУРА

- [1] Шушкевич, Е.В. Эффективное управление системой подачи и распределения воды Московского мегаполиса / Е.В. Шушкевич. – Журнал ВСТ. – 2011. – Т. 1. – С. 24-30.
- [2] Сниток, В.Е. Прогнозирование. Модели, методы, алгоритмы / В.Е. Сниток. – Киев: Маклауг, 2008. – 367 с.
- [3] Lomet, A. Statistical Modeling for Real Domestic Hot Water Consumption Forecasting / A. Lomet, Fr. Suard, D. Chèze, Energy Procedia. – 2015. – Vol. 70. – P. 379-387. DOI: 10.1016/j.egypro.2015.02.138.
- [4] Chen, G. Multiple random forests modelling for urban water consumption forecasting / G. Chen, T. Long, J. Xiong, Y. Bai // Water Resour. Manag. – 2017. – Vol. 31(15). – P. 4715-4729.
- [5] Задачин, В.М. Система прогнозирования водо-потребления города / В.М. Задачин // Систем. Обр. Инф. – 2010. – Т. 7, № 88. – С. 95-98.
- [6] Бююль, А. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей / А. Бююль, П. Цёфель. – СПб: ДиаСофтЮП, 2020. – 608 с.
- [7] Якимов, И.М. Параллельный генетический алгоритм отбора значимых признаков в задаче формирования мероприятий для нефтяных скважин / И.М. Якимов, П.И. Тутубалин, В.А. Суздальцев, И.А. Зарайский, Э.Г. Тахавова // Международная конференция по мягким вычислениям и измерениям. – 2018. – Т. 1. – С. 720-723.
- [8] Мокшин, В.В. Рекурсивный алгоритм построения регрессионных моделей сложных вероятностных объектов / В.В. Мокшин, И.Р. Сайфутдинов, А.П. Кирпичников // Вестник Технологического университета. – 2017. – Т. 20, № 9. – С. 112-116.

Разложение *in vivo* рамановских спектров кожи методом разрешения многомерных кривых

И.А. Матвеева
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
m-irene-a@yandex.ru

О.О. Мякинин
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
myakole@gmail.com

И.А. Братченко
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
iabratchenko@gmail.com

Аннотация—В последние годы для исследования биологических тканей и жидкостей все чаще используется рамановская спектроскопия. Целью исследования является изучение возможностей метода разрешения многомерных кривых для анализа *in vivo* рамановских спектров кожи. Проведено разложение рамановских спектров здоровой кожи и кожи с кератозом, базальноклеточной карциномой, злокачественной меланомой и пигментным невусом. В результате разложения экспериментальных рамановских спектров получены рамановские спектры, соответствующие вкладу оптической системы и компонентов кожи: меланина, белков, липидов, воды и др. Результаты демонстрируют возможность выделения нескольких спектрально близких компонентов с помощью метода разрешения многомерных кривых даже из зашумленных экспериментальных рамановских спектров. Классификация рамановских спектров разных нозологий (злокачественные и доброкачественные новообразования, злокачественная меланома и пигментные новообразования) по вкладу их отдельных компонентов показывает точность классификации около 70%.

Ключевые слова— доброкачественное новообразование, злокачественное новообразование, классификация, метод разрешения многомерных кривых, рамановская спектроскопия.

1. ВВЕДЕНИЕ

При развитии заболевания в организме пациента происходят метаболические изменения, что приводит к изменению биохимического состава биологических тканей и жидкостей. В последние годы для исследования этих изменений все чаще используется спектроскопия рамановского рассеяния [1]. Однако входящие в состав биологической ткани вещества имеют перекрывающиеся рамановские полосы, и из анализируемого рамановского спектра кожной ткани трудно выделить отдельные компоненты. Кроме того, при большинстве используемых сейчас методов анализа рамановских спектров невозможна физическая интерпретация результатов. Целью данного исследования является изучение возможностей метода разрешения многомерных кривых (MCR) [2] для анализа рамановских спектров кожи, зарегистрированных *in vivo* с использованием портативной спектроскопической установки [3].

2. МАТЕРИАЛЫ И МЕТОДЫ

В работе использованы 1000 рамановских спектров: 540 спектров здоровой кожи, 113 кератоза (К), 122 базальноклеточного рака (БСР), 67 злокачественной меланомы (ММ) и 158 пигментного невуса (ПН).

Рамановские спектры кожи регистрировались *in vivo* с помощью портативной спектроскопической установки [3]. Спектры обрезаны в диапазоне от 860 до 920 нм, что соответствует $1114-1874 \text{ cm}^{-1}$ и подвергнуты предварительной обработке с удалением базовой линии и сглаживанием [4].

Для разложения рамановских спектров методом MCR использован протокол *Felten et al.* [5]. Основная идея метода MCR заключается в разложении матрицы рамановских спектров D на более мелкие матрицы C и S^T , где C представляет собой профили концентраций для каждого компонента кожи, а S^T представляет собой матрицу рамановских спектров компонентов. После того, как для матрицы C дана начальная оценка, она итеративно оптимизируется с использованием метода чередующихся наименьших квадратов (ALS) до тех пор, пока не будет достигнута сходимость.

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В результате разложения рамановских спектров кожи методом MCR выделены несколько компонентов, отражающих вклад оптической системы регистрации и некоторых групп веществ (см. Рис. 1).

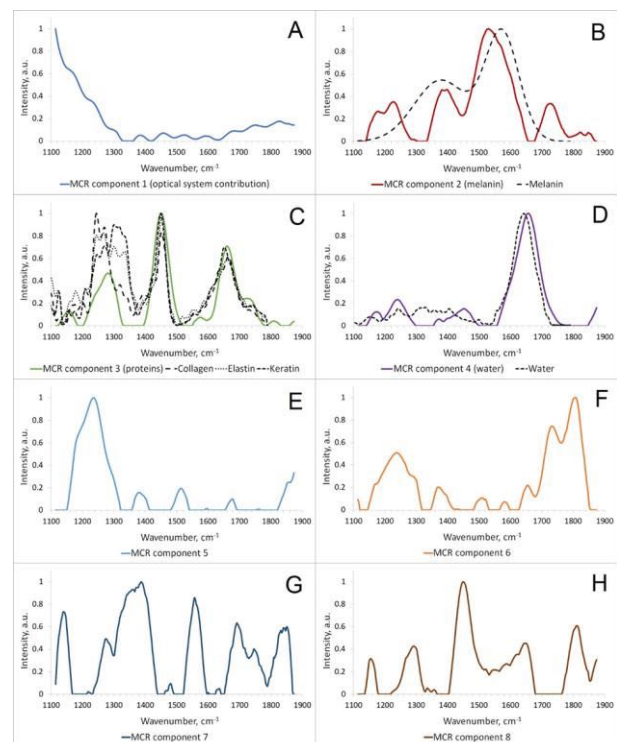


Рис. 1. Рамановские спектры компонентов кожи (сплошные линии – спектры, полученные в данном исследовании, штриховые линии – спектры, полученные в [19])

Компонент 1 (Рис. 1, А) соответствует вкладу оптической системы [6]. В спектре компонента 2 (Рис. 1, В) пики 1390 и 1520 см^{-1} соответствуют меланину [6, 7]. В спектре компонента 3 (Рис. 1, С) пик 1280 см^{-1} – это амид III, 1450 см^{-1} – вклад кератина, коллагена и эластина, а 1660 см^{-1} – вклад коллагена [7]. В спектре компонента 4 (Рис. 1, D) наиболее интенсивный пик 1650 см^{-1} соответствует вкладу воды [7]. Спектр компонента 5 (Рис. 1, E) может соответствовать вкладу липидов, а также натурального увлажняющего фактора [6, 7]. Кроме того, в спектрах компонентов 5 и 6 (Рис. 1, E, F) виден интенсивный пик при 1240 см^{-1} , соответствующий белкам. Спектр компонента 6 (Рис. 1, F) также может соответствовать вкладу липидов [7]. В спектре компонента 7 (Рис. 1, G) можно отметить пики натурального увлажняющего фактора, липидов, меланина и коллагена [6, 7]. В спектре компонента 8 (Рис. 1, H) видны интенсивные пики 1450 и 1650 см^{-1} , которые являются вкладом белков и воды, соответственно.

Далее для некоторых случаев классификации была применена логистическая регрессия (см. Рис. 2).

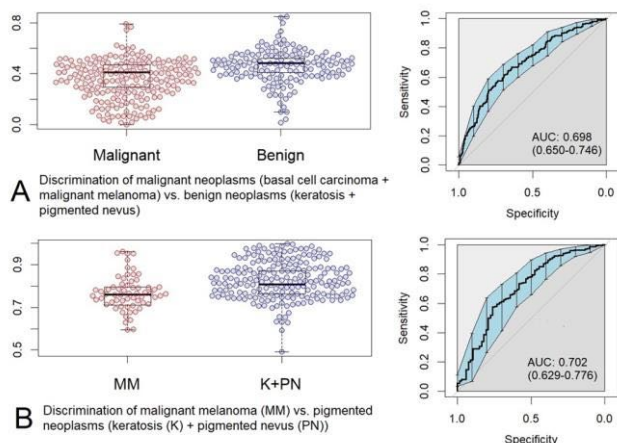


Рис. 2. Диаграммы размаха и ROC-кривые предсказанных значений, полученных в результате логистической регрессии: А – злокачественные новообразования (BCC + MM) и доброкачественные новообразования (K + PN), В – злокачественная меланома (MM) и пигментные новообразования (K + PN)

Как видно из Рис. 2, в случае классификации злокачественных (MM, BCC; $n = 189$) и доброкачественных (K, PN; $n = 271$) новообразований ROC AUC составляет 0,698 (0,650-0,746, 95% CI), а в случае классификации MM ($n = 67$) и пигментных новообразований (K, PN; $n = 271$) ROC AUC составляет 0,702 (0,629–0,776, 95% CI).

В [3] данные обрабатывались с использованием дискриминантного анализа методом частичных наименьших квадратов (PLS-DA), и AUC аналогичных моделей классификации составили 0,75 (0,71-0,79) и 0,61 (0,53-0,69) соответственно. В результате сравнения можно сделать вывод, что PLS-DA (применительно к рамановским спектрам [3]) и анализ методом MCR с последующей логистической регрессией дают примерно одинаковые результаты точности классификации.

4. ЗАКЛЮЧЕНИЕ

Полученные результаты показывают, что анализ рамановских спектров методом MCR позволяет получить новую информацию о биохимическом составе кожной ткани. Такая информация может быть использована в медицинской диагностике и скрининге населения, когда регистрация рамановских спектров производится с низким отношением сигнал/шум.

Помимо первоначальной задачи анализа рамановских с целью их классификации, также выявлена возможность использования метода MCR для выделения паразитного сигнала, связанного со вкладом оптической системы регистрации спектров.

Среди недостатков метода MCR можно выделить сложность разделения спектрально похожих веществ. Поэтому выделение спектров отдельных белков или липидов по-прежнему остается сложной задачей.

Тем не менее, несмотря на некоторые ограничения метода MCR, получен удовлетворительный результат классификации некоторых заболеваний. Дальнейшие исследования планируется посвятить дополнению используемой методики другими методами анализа рамановских спектров, например, с помощью нейронных сетей.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта РФФИ № 21-75-10097.

ЛИТЕРАТУРА

- [1] Popp, J. Modern Raman spectroscopy for biomedical applications: A variety of Raman spectroscopic techniques on the threshold of biomedical applications / J. Popp, C. Krafft, T. Mayerhöfer // *Optik & Photonik*. – 2011. – Vol. 6(4). – P. 24-28.
- [2] De Juan, A. Multivariate Curve Resolution: 50 years addressing the mixture analysis problem—A review / A. de Juan, R. Tauler // *Analytica Chimica Acta*. – 2021. – Vol. 1145. – P. 59-78. DOI: 10.1016/j.aca.2020.10.051.
- [3] Bratchenko, I.A. In vivo diagnosis of skin cancer with a portable Raman spectroscopic device / I.A. Bratchenko, L.A. Bratchenko, A.A. Moryatov, Y.A. Khristoforova, D.N. Artemyev, O.O. Myakinin, A.E. Orlov, S.V. Kozlov, V.P. Zakharov // *Experimental Dermatology*. – 2021. – Vol. 30(5). – P. 652-663. DOI: 10.1111/exd.14301.
- [4] “TPT-cloud” chemometrics toolbox [Электронный ресурс]. – Режим доступа: <https://tptcloud.com/#/> (07.12.2021).
- [5] Felten, J. Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS) / J. Felten, H. Hall, J. Jaumot, R. Tauler, A. De Juan, A. Gorzsás // *Nature protocols*. – 2015. – Vol. 10(2). – P. 217-240. DOI: 10.1038/nprot.2015.008.
- [6] Yakimov, B.P. Blind source separation of molecular components of the human skin in vivo: non-negative matrix factorization of Raman microspectroscopy data / B.P. Yakimov, A.V. Venets, J. Schleusener, V.V. Fadeev, J. Lademann, E.A. Shirshin, M.E. Darwin // *Analyst*. – 2021. – Vol. 146(10). – P. 3185-3196. DOI: 10.1039/d0an02480e.
- [7] Feng, X. Raman active components of skin cancer / X. Feng, A.J. Moy, H.T. Nguyen, J. Zhang, M.C. Fox, K.R. Sebastian, J.S. Reichenberg, M.K. Markey, J.W. Tunnell // *Biomedical Optics Express*. – 2017. – Vol. 8(6). – P. 2835-2850. DOI: 10.1364/BOE.8.002835.

Новые алгоритмы дискретной фильтрации на основе MWGS-ортогонализации для систем с мультипликативными и аддитивными шумами

А.В. Цыганов
Ульяновский государственный
педагогический университет
им. И.Н. Ульянова
Ульяновск, Россия
andrew.tsyganov@gmail.com

Ю.В. Цыганова
Ульяновский государственный
университет
Ульяновск, Россия
tsyganovajv@gmail.com

А.Н. Кувшинова
Ульяновский государственный
педагогический университет
им. И.Н. Ульянова
Ульяновск, Россия
kuvanulspu@yandex.ru

Т.Н. Куренева
Ульяновский государственный
педагогический университет
им. И.Н. Ульянова
Ульяновск, Россия
tankur67@yandex.ru

Аннотация—В работе рассматривается задача дискретной фильтрации источника данных, моделируемого линейной стохастической системой в пространстве состояний с мультипликативными и аддитивными шумами. Предложен новый класс алгоритмов дискретной фильтрации, основанный на модифицированной взвешенной ортогонализации Грама-Шмидта (MWGS-ортогонализации). Подробно описаны подходы к построению, основные характеристики, достоинства и недостатки алгоритмов. Приведен численный пример их применения для решения задачи дискретной фильтрации.

Ключевые слова— стохастические системы с мультипликативными шумами, дискретная фильтрация, MWGS-ортогонализация, блочно-матричные алгоритмы.

1. ВВЕДЕНИЕ

Дискретные линейные стохастические системы с аддитивными и мультипликативными шумами применяются для решения широкого круга практических задач, связанных с обработкой измерительной информации и фильтрацией данных (например, обработка изображений и сигналов, сетевые технологии, финансовая математика, авионавтика и др.).

Разработка численно эффективных алгоритмов дискретной фильтрации для указанных систем является актуальной и важной задачей, решение которой позволяет получить надежные в вычислительном плане алгоритмы. Одним из подходов к решению является применение вычислительной модели алгоритмов фильтрации, основанной на MWGS-ортогонализации.

Целью данной работы является разработка и теоретическое обоснование нового класса алгоритмов ковариационной и информационной фильтрации данных, моделируемых дискретной стохастической системой с мультипликативными и аддитивными шумами, основанных на едином подходе к обработке данных — MWGS-ортогонализации.

2. НОВЫЙ КЛАСС АЛГОРИТМОВ НА ОСНОВЕ MWGS-ОРТОГОНАЛИЗАЦИИ

Рассмотрим дискретную линейную стохастическую систему:

$$\begin{cases} x_k = (F_{k-1} + \tilde{F}_{k-1}\xi_{k-1})x_{k-1} + G_{k-1}w_{k-1}, \\ z_k = (H_k + \tilde{H}_k\zeta_k)x_k + v_k, \quad k = 1, 2, \dots, \end{cases}$$

где w_k и v_k — аддитивные гауссовские шумы, ξ_k и ζ_k — мультипликативные гауссовские шумы.

Стандартные алгоритмы дискретной фильтрации калмановского типа для рассматриваемых систем известны (см, например, [1,2]). В недавней работе [3] предложен общий подход к построению MWGS-алгоритмов для систем с аддитивными шумами. В данной работе мы развиваем полученные в [3] результаты для более широкого класса систем с мультипликативными и аддитивными шумами. При построении алгоритмов были использованы два типа MWGS-ортогонализации: *прямая* процедура, основанная на LDL^T -разложении блочной матрицы A с учетом блочно-диагональной весовой матрицы D_A и *обратная* процедура, основанная на UDU^T -разложении блочной матрицы A с учетом блочно-диагональной весовой матрицы D_A .

Теоретическое обоснование каждого алгоритма основано на доказательстве его алгебраической эквивалентности стандартным алгоритмам ковариационного и информационного типа. В качестве примера в работе приведена схема доказательства алгоритмов eUD-CF-TM и eUD-IF-TM. Также приведены результаты компьютерного моделирования на языке MATLAB.

В таблице 1 приведен список MWGS-алгоритмов и их основные характеристики. Все алгоритмы являются блочно-матричными. Характеристика “Тип” (расширенный/обычный) указывает на способ вычисления оценок вектора состояния (в едином блочном массиве или в отдельном матричном выражении).

3. ПРИМЕР

В качестве примера рассмотрим модель почти равномерного прямолинейного движения объекта с мультипликативными шумами в объекте и измерителе и плохо обусловленной схемой измерений:

$$\begin{cases} x_k = \left(\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xi_{k-1} \right) x_{k-1} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} w_{k-1}, \\ z_k = \left(\begin{bmatrix} 1 & 1 \\ 1 & 1 + \delta \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \zeta_k \right) x_k + v_k, \end{cases}$$

$x_0 \sim N([0, 1]^T, 10I_2)$, $w_k \sim N(0, 10^{-2})$, $v_k \sim N(0, \delta^2 I_2)$, $\xi_k \sim N(0, 10^{-5})$, $\zeta_k \sim N(0, \delta^2)$, где δ — параметр, моделирующий влияние ошибок машинного округления.

ТАБЛИЦА 1. ХАРАКТЕРИСТИКИ MWGS-АЛГОРИТМОВ

Алгоритм	Форма, тип	Разложение	Этапы
(UD/LD)-CF-TM	Ковариационная, обычный	UDU^T/LDL^T	Экстраполяция + фильтрация
(UD/LD)-CF	Ковариационная, обычный	UDU^T/LDL^T	Совмещенный
e(UD/LD)-CF-TM	Ковариационная, расширенный	UDU^T/LDL^T	Экстраполяция + фильтрация
e(UD/LD)-CF	Ковариационная, расширенный	UDU^T/LDL^T	Совмещенный
(UD/LD)-IF-TM	Информационная, обычный	UDU^T/LDL^T	Экстраполяция + фильтрация
(UD/LD)-IF	Информационная, обычный	UDU^T/LDL^T	Совмещенный
e(UD/LD)-IF-TM	Информационная, расширенный	UDU^T/LDL^T	Экстраполяция + фильтрация
e(UD/LD)-IF	Информационная, расширенный	UDU^T/LDL^T	Совмещенный

Покажем на численных экспериментах преимущества предложенных алгоритмов в плане численной устойчивости по отношению к ошибкам машинного округления. Проведем сравнительный анализ эффективности следующих алгоритмов [2]: стандартного фильтра Калмана в ковариационной форме (CF), двух модификаций информационного фильтра (IF₁ и IF₂), а также двух расширенных UD-фильтров (eUD-CF-TM и eUD-IF-TM) для различных значений параметра δ (число измерений $N = 1000$).

В таблице 2 приведены результаты численных экспериментов в системе MATLAB. Здесь $RMSE(x_1)$, $RMSE(x_2)$ — среднеквадратические ошибки оценивания вектора состояния по x_1 и x_2 соответственно, $nRMSE = \|\mathbf{RMSE}(x)\|_2$. Из приведенных результатов видно, что при $\delta = 10^{-6}$ все алгоритмы показывают примерно одинаковые результаты, при $\delta = 10^{-7}$ ошибки обычных информационных фильтров (IF₁ и IF₂) начинают возрастать.

Случай $\delta = 10^{-8}$ соответствует ситуации, когда $\delta^2 < \epsilon_{\text{roundoff}}$, но $\delta > \epsilon_{\text{roundoff}}$, $\epsilon_{\text{roundoff}}$ — машинный эпсилон. В данных условиях фильтры CF, IF₁ и IF₂ начинают расходиться, в то время как фильтры eUD-CF-TM и eUD-IF-TM продолжают стабильно работать.

4. ЗАКЛЮЧЕНИЕ

В работе предложен новый класс алгоритмов дискретной фильтрации для линейных стохастических систем с мультипликативными и аддитивными шумами.

Все алгоритмы основаны на едином подходе к обработке данных — MWGS-ортогонализации. Подробно описаны основные характеристики алгоритмов, их достоинства и недостатки. Теоретическое обоснование предложенных алгоритмов основано на доказательстве их алгебраической эквивалентности существующим стандартными алгоритмам ковариационного и информационного типа. На численном примере показаны преимущества предложенных алгоритмов в сравнении со стандартными в плане организации вычислений и численной устойчивости к ошибкам машинного округления.

ТАБЛИЦА 2. ОШИБКИ RMSE

$\delta = 10^{-6}$			
	RMSE(x ₁)	RMSE(x ₂)	nRMSE
CF	0.0520	0.0520	0.0736
IF ₁	0.0517	0.0517	0.0732
IF ₂	0.0521	0.0521	0.0736
eUD-CF-TM	0.0520	0.0520	0.0736
eUD-IF-TM	0.0520	0.0520	0.0736
$\delta = 10^{-7}$			
	RMSE(x ₁)	RMSE(x ₂)	nRMSE
CF	0.0519	0.0519	0.0734
IF ₁	0.1518	0.1518	0.2146
IF ₂	0.2094	0.2094	0.2961
eUD-CF-TM	0.0520	0.0520	0.0736
eUD-IF-TM	0.0520	0.0520	0.0736
$\delta = 10^{-8}$			
	RMSE(x ₁)	RMSE(x ₂)	nRMSE
CF	NaN	NaN	NaN
IF ₁	Inf	Inf	Inf
IF ₂	NaN	NaN	NaN
eUD-CF-TM	0.0520	0.0520	0.0736
eUD-IF-TM	0.0520	0.0520	0.0736

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда № 22–21–00387, <https://rscf.ru/project/22-21-00387/>.

ЛИТЕРАТУРА

- [1] Wu, Y. Kalman filtering with multiplicative and additive noises / Y. Wu, Q. Zhang, Z. Shen // Proc. of the 12th World Congress on Intelligent Control and Automation (WCICA). – 2016. – P. 483–487. DOI: 10.1109/WCICA.2016.7578352.
- [2] Tsyganov, A.V. UD-based Linear Filtering for Discrete-Time Systems with Multiplicative and Additive Noises / A.V. Tsyganov, J.V. Tsyganova, T.N. Kureneva // Proc. of the 19th European Control Conference. – 2020. – P. 1389–1394. DOI: 10.23919/ECC51009.2020.9143804.
- [3] Tsyganova, J.V. A general approach for designing the MWGS-based information-form Kalman filtering methods / J.V. Tsyganova, M.V. Kulikova, A.V. Tsyganov // European Journal of Control. – 2020. – Vol. 56. – P. 86–97. DOI: 10.1016/j.ejcon.2020.02.001.

Разработка приложения по моделированию локомоторной активности лабораторных ЖИВОТНЫХ

Д.С. Борисов

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
dimaborisov290699@yandex.ru

А.В. Благов

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
alexander.blagov@gmail.com

Аннотация—В данной работе мы представляем разработанный аналитический инструмент для проведения исследования влияния различных факторов на локомоторную активность лабораторного животного. Целью исследования является выявление изменений в основных характеристиках циркадианного ритма после воздействия того или иного вещества. Создан графический интерфейс пользователя с возможностями составления вывода об исследовании и сохранения результатов работы приложения.

Ключевые слова— локомоторная активность, Python, Ломб-Скаргла.

1. ВВЕДЕНИЕ

Исследование влияния того или иного вещества на локомоторную активность живого существа, разумеется, всегда актуально в научной деятельности различных лабораторий.

Очевидно, самой сложной и математически значимой частью исследования является анализ изменения циркадианного ритма живого существа. Среди исследуемых и сравниваемых параметров находятся такие характеристики циркадианного ритма, как:

- а) период,
- б) амплитуда,
- в) акрофаза,
- г) мезор.

Задачей данной работы является определение характеристики циркадианного ритма до и после события для последующего составления вывода об изменениях.

2. СБОР ДАННЫХ

Сбор информации о локомоторной активности каждой крысы выполняется путём использования специальных датчиков. Они устанавливаются на беговых колесах для каждой крысы в клетке. В клетке находится одно животное и одно беговое колесо. В качестве условной единицы для описания локомоторной активности животного была принята мера поворота колеса на 45 градусов вокруг своей оси - бин. Датчики передают компьютеру количество бинов, совершённых каждой крысой за последние пять секунд наблюдений. Спустя продолжительное время с начала эксперимента на компьютере формируются определённые наборы чисел, характеризующие локомоторную активность лабораторных крыс.

Данные с датчиков записываются в текстовые файлы с уникальным именем в виде таблицы как представлено на рисунке 1.

Время, сек.	Бины 5-ти крыс:				
	Крыса 1	Крыса 2	Крыса 3	Крыса 4	Крыса 5
0	0	0	0	3	7
5	0	0	0	3	1
10	0	0	3	9	2
15	0	0	2	1	1
20	0	0	2	0	0
25	0	0	3	0	1
30	1	0	16	0	0
35	0	0	1	0	0

Рис. 1. Файл с данными от датчиков

Первый столбец таблицы является текущим временем проведения эксперимента, выраженным в секундах, а остальные столбцы – количество бинов, совершённых каждой крысой между отсчётами.

3. ВЫЧИСЛЕНИЕ ПЕРИОДА ЧЕРЕЗ ПЕРИОДОГРАММУ ЛОМБА-СКАРГЛА

Периодограмма Ломба-Скаргла, является классическим методом поиска периодичности в данных с нерегулярной выборкой. Он во многом аналогичен более знакомой спектральной плотности мощности Фурье (PSD), часто используемой для обнаружения периодичности в регулярно отбираемых данных [1].

Периодограмма Ломба-Скаргла включает в себя вычисление мощности $P(\omega)$ для данных $\{y_k\}$ предварительно центрированных таким образом, что $\sum_k y_k = 0$:

$$P(\omega) \cong \frac{|\sum_k y_k \cos \omega(t_k - \tau)|^2}{\sum_k \cos^2 \omega(t_k - \tau)} + \frac{|\sum_k y_k \sin \omega(t_k - \tau)|^2}{\sum_k \sin^2 \omega(t_k - \tau)},$$

где ω – рассматриваемая частота; τ - смещение по времени, которое ортогонализует модель и вычисляется через следующую формулу:

$$\tan(2\omega\tau) = \frac{\sum_k \sin(2\omega t_k)}{\sum_k \cos(2\omega t_k)}.$$

Стоит заметить, что для N количества частот ω выполняется по N тригонометрических операций, следовательно, нативный алгоритм Ломба-Скаргла имеет вычислительную сложность $O(N^2)$. Однако в своей статье 1989 года Уильям Пресс и Джордж Рыбичский продемонстрировали хитрый подход, при котором быстрое преобразование Фурье используется на сетке, экстраполированной из исходных данных, так что данная

$O(N^2)$ задача может быть решена за время $O(N \log N)$ [2].

4. АНАЛИЗ РЕЗУЛЬТАТОВ

Оставшиеся три характеристики можно вычислить, обратившись к их определению:

1. Акрофаза является моментом времени с максимальным значением активности крысы за сутки.

2. Мезор представляет собой среднее значение активности животного, выраженное в бинах.

3. Амплитуда, как наибольший пик активности крысы относительно её периода.

Составление вывода на основе результатов работы программы состоит из двух этапов: определение сдвига для каждой характеристики из-за вещества и косинор-анализ – способ аппроксимация суточных кривых синусоидой [3]. Расчётная функция для синусоиды записывается следующим образом:

$$Y(t) = M + A \cos\left(\frac{2\pi t}{\tau} + \varphi\right) + e(t),$$

где M – мезор; A – амплитуда; φ – акрофаза; τ – период; $e(t)$ – случайная компонента, описывающая прочие факторы, не включённые в функцию.

Из-за отсутствия прочих факторов компоненту $e(t)$ можно не учитывать. График аппроксимированных данных для периодограммы Ломба-Скаргла представлен на рисунке 2.

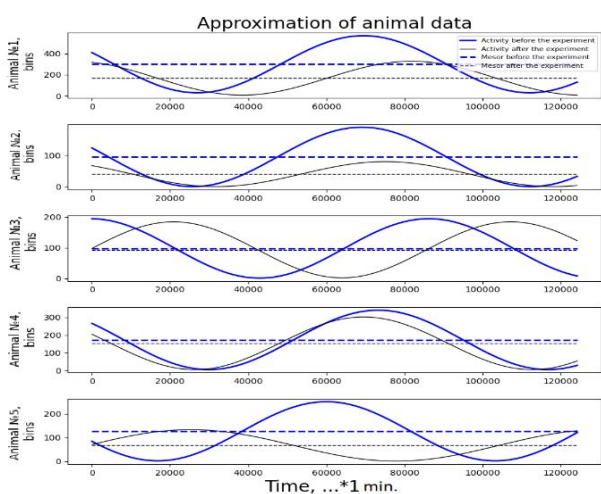


Рис. 2. Косинор-анализ

Вычисленные сдвиги характеристик циркадианного ритма представлены в таблице I.

Анализ проводился на основе данных исследования влияния инсулина на циркадианный ритм крыс [4], предоставленных кафедрой физиологии животного и человека Самарского университета. На основе результатов можно сделать вывод, что инсулин по-разному влияет на циркадианный ритм в зависимости от времени введения вещества организму. Подобный результат также был в собственном исследовании кафедры. Результатом их исследования оказался

наиболее выраженный сдвиг локомоторной активности при введении препарата в момент проецированного суточного цикла равного $ZT = 13$ (анализ сдвига акрофазы выявил значительное опережение на $5,48 \pm 1,98$ часов), незначительное изменение суммарной суточной активности [4].

ТАБЛИЦА 1. СДВИГИ ХАРАКТЕРИСТИК КРЫС

Номер крысы	Периодограмма Ломба-Скаргла			
	Мезор	Амплитуда	Период	Фаза
1	-22	-110	1657,4	23580,0
2	0	-55	-123,6	18780,0
3	1	-6	-308,3	-5625,0
4	0	-19	-529,7	-44302,5
5	-1	-59	18725,4	24595,2

5. ЗАКЛЮЧЕНИЕ

Была проделана работа по исследованию и разработке алгоритмов анализа данных о локомоторной активности крыс для изучения влияния различных веществ на циркадианный ритм животного. Для данной цели был разработан программный инструментальный, позволяющий:

- чтение двух файлов с данными наблюдений;
- их обработка, путём отбрасывания данных с заданного времени до полуночи и их компоновка с определённым шагом;
- сохранение обработанных данных в текстовые и графические файлы;
- анализ характеристик животных с каждого файла и составление вывода об их изменениях с возможностью составления синусоидального графика локомоторной активности животных.

В процессе разработки были использованы такие библиотеки и технологии разработки, как PyQt5, threading и Qt Designer. В процессе нахождения характеристик циркадианного ритма была применена периодограмма Ломба-Скаргла. В дальнейшем планируется расширять функционал и возможности программного комплекса.

ЛИТЕРАТУРА

- [1] Vanderplas, J. Fast Lomb-Scargle Periodograms in Python / J. Vanderplas // Pythonic Perambulations, 2015 [Электронный ресурс]. – Режим доступа: <https://jakevdp.github.io/blog/2015/06/13/lomb-scargle-in-python/> (02.03.2021).
- [2] Press, W.H. Fast algorithm for spectral analysis of unevenly sampled data / W.H. Press, G.B. Rybicki // The Astrophysical Journal. – 1989. – Vol. 338. – P. 277-280.
- [3] Борисов, Д.С. Разработка инструментария по обработке и анализу данных наблюдений за активностью лабораторных животных / Д.С. Борисов, А.В. Благов, А.Н. Инюшкин // Материалы Международной конференции и молодёжной школы «Информационные технологии и нанотехнологии» (ИТНТ-2020). – 2020. – С. 112-117.
- [4] Инюшкин, А.Н. Влияние инсулина на циркадианный ритм произвольной локомоторной активности крыс / А.Н. Инюшкин // Российский физиологический журнал им. ИМ Сеченова. – 2011. – Т. 97, № 2. – С. 678-689.

Оценка качества интерфейса на основе анализа временных рядов результатов обучения пользователей

В.К. Григорьев
МИРЭА-Российский технологический университет
Москва, Россия
grigoriev@mirea.ru

М.А. Овчинников
МИРЭА-Российский технологический университет
Москва, Россия
ovchinnikov_m@mirea.ru

Аннотация—В настоящей работе рассматривается побочный эффект опережающего деятельного обучения пользователей в виде пошаговых рядов времен, полученных при выполнении сценария. Дается структура времени шага в виде полного, ментального и механического времен. Выдвигается гипотеза о возможности замены механического времени по GOMS на минимальное время выполнения шага, полученного на основе статистической обработки временных рядов выполнения сценария. Эта гипотеза подтверждается экспериментальным путём на основе опережающего обучения. Данный результат может быть использован для периодической оценки качества интерфейса в процессе AGILE разработки программного продукта.

Ключевые слова— качество интерфейса, временной ряд исполнения шагов, AGILE-разработка, опережающее обучение.

1. ВВЕДЕНИЕ

Эффективность использования программных продуктов напрямую зависит от качества их интерфейсов и уровня подготовки пользователя [1]. Причём, качество в данном случае определяется не только отсутствием ошибок в интерфейсе, но и его удобством использования массовыми профессиональными пользователями в соответствии современными гайдлайнами [2]. В работе [3] описывается эффективный метод опережающей подготовки МПП с помощью обучающих программ, создаваемых на основе инструментального комплекса «Построитель Тьюторов».

В последнее время широкое распространение при создании новых программных продуктов получил AGILE подход. Его важной особенностью выступает короткий период времени, необходимый для получения некоторого промежуточного экземпляра разрабатываемого программного продукта и его анализ в конце каждого этапа разработки в плотном взаимодействии заказчика и разработчика. AGILE подход, наиболее эффективен при разработке специализированных программных продуктов, ориентированных на решение конкретных задач пользователей [4]. Наличие конкретных задач пользователей позволяют проектировать интерфейс на основе сценария работы пользователя, периодически его корректировать на основе оценки качества интерфейса и при уточнении заказчиком сценария работы пользователя [5]. Оценку интерфейса промежуточного экземпляра программного продукта на каждом этапе разработки можно проводить при помощи инструментального комплекса «Построитель Тьюторов» [6], путем обучения части потенциальных пользователей.

2. МЕТОДИКА ВЫЧИСЛЕНИЯ РЯДА МЕНТАЛЬНЫХ ВРЕМЁН

Действительно, в качестве побочного результата обучения с помощью инструментального комплекса «Построитель Тьюторов» получаем последовательность данных о времени выполнения элементарного шага. Здесь, под элементарным шагом понимается выполнение какого-либо действия пользователем в программном продукте: нажатие кнопки на клавиатуре, мыши, использование колесика или перетаскивание с помощью мыши, приводящие к переходу от одного окна к другому. Таким образом получаем две последовательности – времени выполнения шагов и информации об ошибочности или правильности действий пользователей. Анализ этих двух последовательностей может быть использован для оценки качества интерфейса экземпляра продукта, полученного в конце каждого этапа [7]. Действительно, время выполнения работы во многом характеризует качество интерфейса программного продукта, в предположении о высокой квалификации пользователей в своей предметной области.

Время шага можно рассматривать как сумму ментального ($T_i^{\text{мент.}}$) и механического ($T_i^{\text{мех.}}$) времен (1).

$$T_i = T_i^{\text{мент.}} + T_i^{\text{мех.}} \quad (1)$$

Для оценки качества обучения был проведен эксперимент по обучению 30 человек. Обучающая программа выполнялась три раза. Исследовалось изменение ментального времени, как оценка качества обучения. В качестве побочного результата обучения записывалось время исполнения каждого шага сценария. Сценарий состоял из 49 шагов.

Анализ побочных результатов проведённого эксперимента, характеризующих качество интерфейса через время выполнения отдельных шагов, (рис. 1) позволяет сделать вывод о том, что в основном время выполнения характеризуется ментальным временем.

Таким образом, оценку качества интерфейса мы можем проводить на основе анализа ментальных временных рядов, построенных по формуле (2).

$$T_i^{\text{мент.}} = T_i - T_i^{\text{мех.}} \quad (2)$$

Для реализации формулы (2) необходимо получить механическое время шага.



Рис. 1. Сравнительный анализ временного ряда выполнения шагов и механического времени по GOMS

Средняя оценка механического времени выполнения, не учитывающая особенности конкретных пользователей может быть построена на основе метода KLM-GOMS. Однако, при разработке программного обеспечения для конкретного класса задач и конкретных пользователей, уже прошедших специальный отбор для решения задач данной области, целесообразно ориентироваться не на средние показатели (GOMS), которые не учитывают особенности данного класса пользователей, а использовать экспериментально полученные данные пользователей в процессе опережающего обучения на каждом этапе SCRUM цикла. Для этого случая выдвигается гипотеза, что в качестве механического времени можно использовать минимальное время, имеющееся на данном элементарном шаге (3).

$T_i^{mex.} = \min_j T_{ij}$, где j – от 1 до количества пользователей (3)

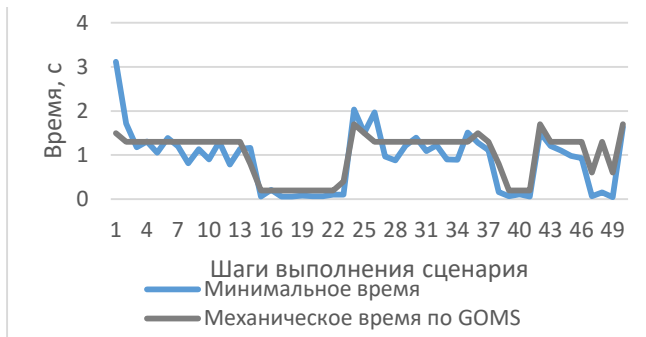


Рис. 2. Сравнительный анализ временного ряда выполнения шагов и механического времени по GOMS

Результаты проведенных экспериментов по сравнению значений $T_i^{mex.}$, построенного на основе метода GOMS, и $T_i^{mex.}$, построенного на основе обработки временных рядов обучения пользователей подтверждают гипотезу о возможности использования (3) для построения механического времени (рис. 2).

Действительно, построение временного ряда по GOMS при каждом изменении элементов сценария требует ручных вычислений, что занимает в условиях AGILE подхода избыточное время. Таким образом, подтверждается целесообразность замены временного ряда по GOMS на автоматически вычисляемый временной ряд.

Как уже отмечалось, важнейшим параметром, который определяет работу команды метод SCRUM, является время. Выполним оценку времени, которое затрачивается на создание обучающей программы и опережающего обучения пользователей (для каждого уточненного программного продукта на каждом цикле). Обучающая программа создается автоматически на основе сценария работы пользователя. Сценарий представлен в начале работы, в частности, для проектирования интерфейса. На каждом цикле SCRUM заказчики-пользователи уточняют сценарий и, в соответствии с этим, модифицируется интерфейс. В конце цикла желательно получить оценку качества этого интерфейса, так как экспертные оценки обладают существенными недостатками: занимают много времени и не объективны, то целесообразно использовать для оценки качества численные объективные значения. Это целесообразно делать на основе временных рядов, полученных как побочный эффект опережающего обучения пользователей. Из работы [6], мы видим, что сам процесс автоматизированного создания обучающей программы занимает незначительное время. При использовании метода SCRUM для создания программного продукта одной из важных характеристик является возможность простого и удобного контакта с потенциальными пользователями заказчика. Таким образом, процесс опережающего обучения может быть проведен в рабочее время пользователей, практически без серьезного отвлечения их от работы.

3. ЗАКЛЮЧЕНИЕ

Предложенная авторами и экспериментально подтвержденная гипотеза о возможности замены ряда механического времени по GOMS на ряд минимальных значений времени выполнения шагов позволяет сделать вывод о целесообразности использования обучающих программ для объективной численной оценки качества интерфейса программного продукта, разрабатываемого на основе метода SCRUM.

ЛИТЕРАТУРА

- [1] Цветков, В.Я Информационное взаимодействие в человеко-машинной системе / Цветков, В.Я // Образовательные ресурсы и технологии. – 2021. – № 3. – С. 88-96.
- [2] Senevirathne, G. Impact of E-Learning System User Interface Design on User Satisfaction / G. Senevirathne, K. Manathunga // 9th Edition of IEEE Region 10 Humanitarian Technology Conference. – Bangalore. – 2021. – Vol. R10-HTC. DOI: 10.1109/R10-HTC53172. 2021.9641570.
- [3] Григорьев, В.К. Обобщенный интерфейс системы для обучения массовых профессиональных пользователей / В.К. Григорьев // Открытое образование. – 2010. – № 1. – С. 11-18.
- [4] Pichler, R. Agile Product Management with Scrum: creating products that customers love / R. Pichler // Addison-Wesley, 2010. – 155 p.
- [5] Кудж, С.А. Поддержка жизненного цикла программных компонент / С.А. Кудж, В.Я. Цветков, И.Е. Рогов // Russian Technological Journal. – 2020. – Т. 8, № 5. – С. 19-33. DOI: 10.32362/2500-316X-2020-8-5-19-33.
- [6] Григорьев, В.К. Оценка качества пользовательского интерфейса на основе ментального времени выполнения пользовательских задач предметной области / В.К. Григорьев, А.С. Илющечкин, М.А. Овчинников // Russian Technological Journal. – 2019. – Т. 7, № 1. – С. 38-47. DOI: 10.32362/2500-316X-2019-7-1-38-47.
- [7] Басок, Б.М. Некоторые формализованные подходы к оценке удобства интерфейса пользователя веб-приложений / Б.М. Басок, С.Л. Френкель // Russian Technological Journal. – 2021. – Т. 9, № 2. – С. 7-21. DOI: 10.32362/2500-316X-2021-9-2-7-21.

Исследование и разработка алгоритма минимизации звуковой петли в системах громкоговорящей связи

А.А. Колпаков
Владимирский государственный университет
Владимир, Россия
desT.087@gmail.com

Д.В. Бейлекчи
Владимирский государственный университет
Владимир, Россия
dmibeil@yandex.ru

А.Ю. Проскураков
Владимирский государственный университет
Владимир, Россия
alexander.prosk.murom@gmail.com

А.А. Белов
Владимирский государственный университет
Владимир, Россия
aleks.murom@mail.ru

Аннотация—Работа посвящена разработке методов минимизации звуковой петли в телекоммуникационных системах громкоговорящей связи. Результатами исследования является алгоритм минимизации звуковой петли в системах громкоговорящей связи и параметры его практического применения.

Ключевые слова— телекоммуникационные системы, диспетчерская связь, громкоговорящая связь, звуковая петля, обработка звука.

1. ВВЕДЕНИЕ

При проектировании телекоммуникационных систем громкоговорящей связи основным критерием качества является отсутствие помех при передаче и приеме голосовых сообщений. Одной из таких помех является эффект звуковой петли, когда микрофон и громкоговоритель взаимодействуют друг с другом [1-2]. Такую проблему можно решить двумя способами: аппаратными решениями или программными алгоритмами. Аппаратные решения не всегда являются экономически оправданными, поэтому в данной работе рассматривается решение проблемы звуковой петли программным способом.

Основной сценарий использования устройства громкоговорящей связи выглядит следующим образом: оператор находится на своем рабочем месте, обычно за столом, устройство связи расположено на столе в непосредственной близости от оператора, но не ближе 30 см. Оператор произносит фразы во время сеанса связи в нормальном темпе и с нормальной громкостью, не повышая голоса. Принятый сигнал воспроизводится обратно на громкоговоритель. Фоновый шум не превышает 60 дБ [3-4].

Для обеспечения приемлемого уровня разборчивости речи требуется обеспечить следующие параметров качества звукового сигнала:

- отсутствие эффекта звуковой петли;
- при устранении эффекта звуковой петли исключить обрывание речевых фраз;
- уверенный прием речевого сигнала с расстояния до 50 см от микрофона [2].

2. РАЗРАБОТКА МОДУЛЯ ПОДАВЛЕНИЯ ЗВУКОВОЙ ПЕТЛИ

Для подавления эффекта звуковой петли в режиме дуплекс был разработан алгоритм и программный модуль на языке с++, реализующий фильтрацию входных и выходных сигналов с учетом текущего уровня сигнала, обеспечивающий приглушение сигнала входного или выходного сигнала на одном из абонентских устройств при встречном разговоре.

Схема модуля подавления эффекта звуковой петли в режиме связи дуплекс приведена на рисунке 1.

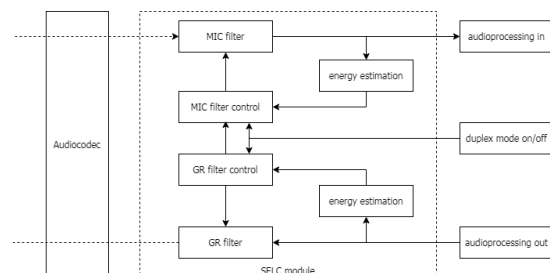


Рис. 1. Структурная схема программного модуля подавления эффекта звуковой петли

В обычном режиме входные и выходные аудиосигналы обрабатываются независимо. Однако расположение динамика и микрофона в непосредственной близости друг от друга вызывает появление эффекта петли акустической обратной связи. Решение этой проблемы может быть обеспечено несколькими методами. Во-первых, можно установить шумозащитный микрофон, который будет фильтровать звук из громкоговорителя. Однако такой подход может значительно увеличить стоимость устройства и сложность его компоновки. Во-вторых, можно отключить громкоговоритель при обнаружении сигнала на микрофоне. Это делает устройство более простым и универсальным, но усложняет программное обеспечение. В данной работе рассматривается второй подход.

Так как регулировка чувствительности аппаратно выполняется на входе до применения к входному сигналу алгоритма подавления звуковой петли, для исключения обрывания речевых фраз при заданном уровне

чувствительности, необходимо выполнять автоматическое задание параметров порогового уровня детектора сигнала с микрофона.

Для оценки требуемого порогового уровня проведены экспериментальные исследования параметров принимаемого сигнала на расстоянии 45 см с разными уровнями чувствительности. Оценка проводилась при измерении фонового шума, а также источника тестовых сигналов.

Оценка энергии входного сигнала в цифровых единицах выполняется по формуле:

$$E_{est} = \frac{\sum_{i=0}^{SAI_BUF_N} s_{in}^2(i)}{256} \quad (1)$$

где $s_{in}(i)$ – отсчеты входного сигнала в формате int16,

SAI_BUF_N – количество отчетов в буфере интерфейса аудиокодека, равно 160 и соответствующее отрезку сигнала 20мс.

При этом получаемое значение E_{est} пропорционально энергии сигнала на входе с коэффициентом 160/256 [3, 4].

3. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ АЛГОРИТМА И РАСЧЕТ УРОВНЕЙ ДЕТЕКТОРА

В ходе экспериментальных исследований, было установлено, что для обеспечения работы алгоритма без обрывания речевых фраз при источнике звука на расстоянии до 50 см при настройках чувствительности по умолчанию при 0 дБ пороговое значение L_{thmic0} должно соответствовать E_{est} при $E_{est} = 400000$ [5].

Пороговые уровни детектора для остальных уровней чувствительности микрофона могут быть рассчитаны по формуле:

$$L_{thmicL} = L_{thmic0} \cdot 10^{\frac{L-L_{st}}{10}} \quad (2)$$

где L – требуемый уровень, L_{st} – шаг изменения уровня в дБ, L_{thmic0} – пороговое значение для уровня 0 дБ.

Были рассчитаны уровни порога детектора с учетом текущего шага регулировки чувствительности $L_{st} = 1,5$ дБ для требуемых уровней [6]:

ТАБЛИЦА I. УРОВНИ ПОРОГА ДЕТЕКТОРА С УЧЕТОМ ТЕКУЩЕГО ШАГА РЕГУЛИРОВКИ ЧУВСТВИТЕЛЬНОСТИ

Уровень	Усиление, дБ	Порог
0	0	400000
-1	-1,5	283200
-2	-3	200505
-3	-4,5	141957
-4	-6	100506
-5	-7,5	71158
-6	-9	50380
+1	1,5	35669
+2	3	25253
+3	4,5	564971
+4	6	797982
+5	7,5	1127094
+6	9	1591940

4. ВЫВОДЫ

В результате проведенных исследований были определены уровни усиления с исключением перегрузок в аналоговой части аппаратного обеспечения по входу и выходу и обеспечивающие прием сигнала на микрофон с расстояния до 50 см, разработан программный код для устранения эффекта звуковой петли в режиме дуплекса и рассчитаны параметры алгоритма для исключения обрывания речевых фраз, что в совокупности обеспечивает требуемые параметры качества звукового сигнала.

ЛИТЕРАТУРА

- [1] Директоров, Н.Ф. Современные системы внутрикорабельной связи. / Н.Ф. Директоров, А.А. Катанович. – СПб.: Судостроение, 2001. – 256 с.
- [2] Кондратьев, К.В. Ограничительные условия реализации метода адаптивной компенсации акустической обратной связи / К.В. Кондратьев, О.В. Непомнящий, А.Ф. Шишкина, В.Н. Сергеевич // Инженерная физика. – 2016. – № 11. – С. 32-38.
- [3] Häsler E. Topics in Acoustic Echo and Noise Control: Selected Methods for the Cancellation of Acoustical Echoes, the Reduction of Background Noise, and Speech Processing / E. Häsler, G. Schmidt. – 2006. – DOI: 10.1007/3-540-33213-8.
- [4] Громоздин, В.В. Имитация внешних акустических шумов при испытаниях качества громкоговорящей связи / В.В. Громоздин, Е.А. Надобенко, Т.В. Панина // 28-я Международная Крымская Конференция "Свч-Техника и Телекоммуникационные Технологии". Материалы конференции. – 2018. – С. 459-465.
- [5] Cohen, I. Speech Processing in Modern Communication-Challenges and Perspectives / I. Cohen, J. Benesty, Sh. Gannot // Springer-Verlag, 2010. – DOI: 10.1007/978-3-642-11130-3.
- [6] Кропотов, Ю.А. Исследование и разработка систем громкоговорящей связи в условиях воздействия акустических помех / Ю.А. Кропотов, А.А. Быков. – Муром: МИ ВлГУ, 2011. – 107 с.
- [7] Кондратьев, К.В. Особенности адаптации фильтров акустической обратной связи в микрофонных блоках систем громкоговорящей связи / К.В. Кондратьев, О.В. Непомнящий, В.Н. Сергеевич // Современные проблемы науки и образования в техническом вузе. Материалы Всероссийской научно-практической конференции. – Уфимский государственный авиационный технический университет, 2013. – Р. 87-91.
- [8] Ермолаев, В.А. Обработка акустических сигналов методами локального анализа в телекоммуникационных системах / В.А. Ермолаев, Ю.А. Кропотов, Д.В. Бейлекчи // Радиотехнические и телекоммуникационные системы. – 2015. – № 1(17). – С. 49-56.
- [9] Kropotov, Yu.A. On the transmission of asynchronous data streams over packet switched networks with random multiple access / Yu.A. Kropotov, A.A. Kolpakov // International Journal of Computer Networks & Communications (IJCNC). – 2018. – Vol. 10(1). – P. 107-117. DOI: 10.5121/ijcnc.2018.10108.
- [10] Rabiner, L. Theory and application of digital signal processing / L. Rabiner, B. Gould. – Moscow, Russia: WORLD, 1978. – 848 p.
- [11] Kropotov, Yu.A. Mathematical models of telecommunication systems with acoustic feedback / Yu.A. Kropotov, A.A. Belov, A.Y. Proskuryakov, A.A. Kolpakov // International Multi-conference on Industrial Engineering and Modern Technologies. – 2019. – P. 8934203.

Спектральный анализ стабิโลграфических сигналов методами Фурье и Гильберта – Хуанга

М.В. Дакинова
Казанский (Приволжский)
федеральный университет
Казань, Россия
MVDakinova@stud.kpfu.ru

Л.М. Бикчентаева
Казанский (Приволжский)
федеральный университет
Казань, Россия
leysanbm@mail.ru

О.А. Саченков
Казанский (Приволжский)
федеральный университет
Казань, Россия
OASachenkov@kpfu.ru

Т.В. Балтина
Казанский (Приволжский)
федеральный университет
Казань, Россия
tvbaltina@kpfu.ru

Г.Г. Яфарова
Казанский (Приволжский)
федеральный университет
Казань, Россия
Gusel.Yafarova@kpfu.ru

Аннотация—Спектральный анализ представляет собой важную часть при цифровой обработке сигналов с целью получить информацию о свойствах, содержащихся в них. Часто используемый метод преобразования Фурье, ограничен при обработке нестационарных и нелинейных сигналов. С каждым годом все больше людей применяют спектральный анализ преобразования Гильберта – Хуанга, в основе которого лежат спектр Гильберта и эмпирическая модовая декомпозиция. Рассмотрим данный метод, применив его к данным стабิโลграммы, служащим оценкой человеком функции равновесия и механизмов поддержания вертикального положения.

Ключевые слова— преобразование Гильберта – Хуанга, эмпирическая модовая декомпозиция, преобразование Фурье, стабิโลграмма, спектральный анализ.

1. ВВЕДЕНИЕ

На данный момент в пакетах прикладных программ анализ стабิโลграмм базируется на классическом методе преобразования Фурье, однако существуют некоторые значительные ограничения спектрального Фурье-анализа: исследуемая система должна быть линейной, а анализируемые данные – стационарными [2]. Для точного определения характеристик стабิโลграммы, воспользуемся преобразованием Гильберта – Хуанга. Данный метод основан на построении спектра Гильберта с использованием эмпирической модовой декомпозиции (ЭМД). ЭМД предложен Норденом Хуангом в 1998 году и представляет собой разложение сложно структурных сигналов на эмпирические моды (ЭМ). Ключевая идея метода состоит в построении базисных функций сигнала, формирующиеся адаптивно из входных данных. Главным свойством ЭМД является его применимость для любых исходных данных.

2. МЕТОДИКА

Цель исследования - спектральный анализ стабิโลграфических сигналов методами Фурье и Гильберта - Хуанга. Задача заключается в декомпозиции сигнала стабิโลграммы на ЭМ и применении их к спектру Гильберта, а также получении параметров спектрограммы.

К механизмам поддержания равновесия у человека относятся: вестибулярная, зрительная и сенсорные системы. В процессе удержания человеком равновесия и поддержания вертикального положения возникают микро

колебательные движения, которые фактически незаметны визуально [1]. Эти колебательные движения отражены на стабิโลграмме и имеют характерные физические свойства.

Чтобы провести частотно – временной анализ стабิโลграммы, строится спектр Гильберта, примененный к ЭМ. Для его построения необходимо знать мгновенную энергию и частоту, которые рассчитываются по следующим формулам:

$$e = |A(t)|^2, \quad (1)$$

$$\omega(t) = d\varphi(t)/dt. \quad (2)$$

Суть ЭМД состоит в том, чтобы представить исходный сигнал суммой компонент, каждая из которых характеризуется колебательным процессом с определенной частотой, присущей исходным данным, и остатком, который показывает тренд, содержащийся в исходных данных.

$$f(t) = \sum imf f_i + res. \quad (3)$$

Подробный алгоритм ЭМД и получение формул (1) – (3) приведено в [5].

Реализация ЭМД и спектра Гильберта производится в пакете прикладных программ MATLAB. Для стандартного способа анализа спектрограммы преобразованием Фурье рассчитаны классические стабิโลграфические показатели [3].

3. ЗАКЛЮЧЕНИЕ

Для реализации поставленной задачи, были обработаны данные стабิโลграммы, когда человек стоит в течение минуты без дополнительных условий, и рассматриваются следующие диапазоны спектров: 0 – 0.1 Гц, 0.1 – 0.5 Гц, 0.5 – 2 Гц, 2 – 4 Гц.

На Рис. 1 представлена спектрограмма, построенная с помощью преобразования Фурье, в которой по диапазонам считают следующие параметры: средние и максимальные амплитуды, процентный вклад мощности, абсолютное значение мощности спектра мощности, а также 60% мощности во всем рассматриваемом диапазоне. Рассчитанные параметры для данной стабิโลграммы представлены в Таблице I.

Таблица 1. ПАРАМЕТРЫ СПЕКТРОГРАММЫ

ПАРАМЕТРЫ																
МАКС-НАЯ АМПЛ-ДА (ММ)				СРЕДНЯЯ АМПЛ-ДА (ММ)				60% (Гц)	МОЩНОСТЬ (%)				АБСОЛ-НАЯ МОЩНОСТЬ (ММ ² /СЕК)			
1	2	3	4	1	2	3	4		1	2	3	4	1	2	3	4
0,6823	0,4451	0,1573	0,0123	0,3673	0,1641	0,0304	0,0021	0,403	24	42	29	5	0,0168	0,0156	0,0027	0,000038

На Рис. 2 представлен гильбертов спектр с наиболее выраженными пятью участками. Из него можно проанализировать частотно – временную зависимость, а также мгновенную энергию, изменяющуюся по ходу эксперимента. Из Рис. 2 и Рис. 3 можно сделать вывод о том, как происходили колебательные процессы в процессе стояния человека. Наиболее высокие частоты с выраженной энергией приходятся на первые секунды, что объясняется установлением равновесия на старте эксперимента. Заметно, что с середины эксперимента снова начинает возрастать значение энергии. Со временем баланс начинает нарушаться, так как возникает незначительное утомление, и амплитуда колебаний увеличивается. Для балансировки тела начинает использоваться голеностопная стратегия, помогающая восстановить общий центр масс движениями в голеностопных суставах, которые активируют лодыжки, колени, бедра и их мышцы [4].

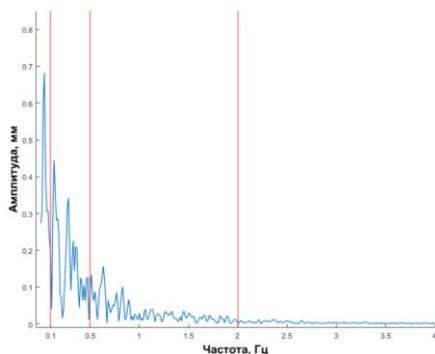


Рис.1. Амплитудный спектр преобразования Фурье с выделенными на частотной области рассматриваемыми диапазонами

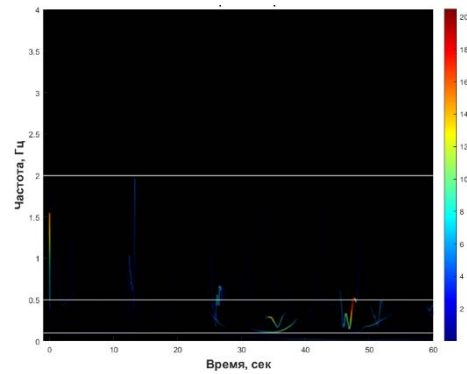


Рис.2. Спектр Гильберта с выделенными на частотной области рассматриваемыми диапазонами

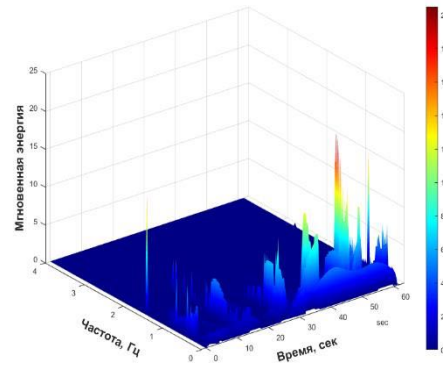


Рис.3. Трехмерное представление зависимостей частота – время – энергия, полученных при преобразовании Гильберта – Хуанга

4. БЛАГОДАРНОСТИ

Работа выполнена за счет средств субсидии, выделенной Казанскому федеральному университету для выполнения государственного задания в сфере научной деятельности, проект № 0671-2020-0059.

ЛИТЕРАТУРА

- [1] Доценко, В.И. Современные алгоритмы стабилметрической диагностики постуральных нарушений в клинической практике / В.И. Доценко, В.И. Усачев, С.В. Морозова, М.А. Скедина // МС. – 2017.
- [2] Микулович, А.В. Анализ нестационарных сигналов на основе преобразования Гильберта–Хуанга / А.В. Микулович, Ш.Ч. Кан, В.И. Микулович // Информатика. – 2010. – № 2. – С. 25-35.
- [3] Руководство пользователя «Стабилан – 01 – 2». – Таганрог: ЗАО «ОКБ РИТМ».
- [4] Скворцов, Д.В. Стабилметрическое исследование: краткое руководство / Д.В. Скворцов. – М.: Маска, 2010. – 176 с.
- [5] Huang, N.E. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis / N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu // Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences. – 1971. – Vol. 454. – P. 903-995.

Метод оценки движения камеры на основе компланарности, инвариантный к глубине сцены

Е.В. Гошин

Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
goshine@ssau.ru

Abstract—В настоящей работе предлагается метод оценки параметров движения камеры по изображениям, полученным с этой камеры. Предложенный подход может применяться к трёхмерным сценам независимо от их глубины и позволяет избежать ограничений, накладываемых на набор исходных данных при использовании методов на основе существенной матрицы и матрицы проективного преобразования.

Ключевые слова— оценка движения камеры, эпиполярная геометрия, проективная геометрия, компланарность.

1. ВВЕДЕНИЕ

В настоящей работе рассматривается задача монокулярной визуальной одометрии [1]. Исследуемая задача состоит в оценке параметров движения камеры [2], характеризующих ориентацию и положение этой камеры в пространстве в определенные моменты времени в процессе движения, по изображениям, регистрируемыми этой камерой.

Исходными данными (при отсутствии информации о перемещении устройства) являются изображения, полученные с разных ракурсов. Традиционно задача оценки параметров поворота и сдвига сводится к определению так называемой существенной матрицы и последующему вычислению искомым параметров [3]. Однако этот подход имеет ряд недостатков. При нахождении исходных точек в одной плоскости вычисление существенной матрицы может осуществляться с ошибками, что, в свою очередь, приводит к ошибкам в определении параметров движения.

Настоящая работа исследует решение принципиальной проблемы поиска универсального способа оценки параметров движения, инвариантного к глубине сцены.

2. МАТЕМАТИЧЕСКАЯ ОСНОВА АЛГОРИТМА

В основе предлагаемого подхода лежит тот факт, что даже при отсутствии информации о расстоянии до точки в трехмерном пространстве можно установить однозначную связь между положением точки после поворота, положением соответствующей точки на втором изображении и вектором сдвига.

Рассмотрим следующую постановку задачи. Пусть задано множество точек на предыдущем и текущем кадрах $\{p_i\} = \{(x_i, y_i, 1)\}$ и $\{p'_i\} = \{(x'_i, y'_i, 1)\}$, полученных в результате проецирования точек $\{M_i\} = \{(X_i, Y_i, Z_i)\}$ и преобразованных точек $\{M'_i\} = \{(X'_i, Y'_i, Z'_i)\} = \{R(X_i, Y_i, Z_i) + t\}$, соответственно.

Необходимо вычислить оценку параметров движения R и t .

Введем обозначения

$$x_i^R = R_{11}x_i + R_{12}y_i + R_{13},$$

$$y_i^R = R_{21}x_i + R_{22}y_i + R_{23},$$

$$z_i^R = R_{31}x_i + R_{32}y_i + R_{33}.$$

В предыдущей работе [4] было показано, что прямая на плоскости, проходящая через точки $\left(\frac{x_i^R}{z_i^R}, \frac{y_i^R}{z_i^R}\right)$ и (x'_i, y'_i) , также проходит через точку $\left(\frac{t_x}{t_z}, \frac{t_y}{t_z}\right)$. Это требование можно записать в форме смешанного произведения:

$$\left(\left(\frac{x_i^R}{z_i^R}, \frac{y_i^R}{z_i^R}, 1 \right) \times (x'_i, y'_i, 1) \right) \cdot \left(\frac{t_x}{t_z}, \frac{t_y}{t_z}, 1 \right) = 0.$$

Так как умножение на константу не меняет уравнения, то и следующее выражение будет верным:

$$\left((x_i^R, y_i^R, z_i^R) \times (x'_i, y'_i, 1) \right) \cdot (t_x, t_y, t_z) = 0. \quad (1)$$

Отметим, что вектор t в выражении (1) не зависит ни от исходных точек, ни от параметров поворота, поэтому это выражение может быть декомпозировано.

Определяем векторы v_i как:

$$v_i = (x_i^R, y_i^R, z_i^R) \times (x'_i, y'_i, 1)$$

и матрицу, состоящую из этих векторов-строк как:

$$V(R) = \begin{pmatrix} v_1 \\ \vdots \end{pmatrix}.$$

Полученная матрица $V(R)$ имеет размерность $n \times 3$, где n — количество пар исходных соответствующих точек.

Очевидно, что произведение каждой строки матрицы $V(R)$ на вектор t будет равно нулю и, следовательно, t будет нуль-вектором полученной матрицы

3. АЛГОРИТМ

В общем случае вектор t предполагается ненулевым. Поскольку матрица имеет ненулевой нуль-вектор, ее ранг должен быть равен 2 или меньше. Другими словами, векторы матрицы v_i должны быть компланарными.

Очевидно, что в реальных условиях при наличии шума в исходных данных компланарность невозможна. Поэтому в качестве критерия соответствия параметра вращения исходным данным вместо этого оценим мультиколлинеарность строк матрицы $V(R)$.

В качестве критерия может быть использован любой из известных критериев мультиколлинеарности: минимальное собственное значение матрицы $V(R)$, число обусловленности матрицы $V(R)$ или определитель матрицы Грама $V^T(R)V(R)$.

После определения параметров поворота путем оптимизации выбранного критерия $J(\mathbf{R})$

$$\hat{\mathbf{R}} = \underset{\mathbf{R}}{\operatorname{argmax}} J(\mathbf{R})$$

значение вектора сдвига \mathbf{t} с точностью до параметра масштаба можно определить путем решения переопределенной однородной системы нелинейных уравнений, исключив тривиальное решение:

$$V(\hat{\mathbf{R}}) \cdot \hat{\mathbf{t}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Можно показать, что для точек, не лежащих на плоскости, единственным значением, удовлетворяющим критерию, является $\hat{\mathbf{R}} = \mathbf{R}$ и $\hat{\mathbf{t}} = \lambda \mathbf{t}$.

4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

Предлагаемый подход был исследован на синтетических данных. Эксперимент заключался в оценке параметров камеры с использованием декомпозиции существенной матрицы и предложенного подхода. Мы использовали угол между точным вектором сдвига \mathbf{t}_{true} и оценкой $\hat{\mathbf{t}}$ как меру ошибки оценки сдвига

$$\epsilon_t = \frac{\mathbf{t}_{true} \cdot \hat{\mathbf{t}}}{\|\mathbf{t}_{true}\| \cdot \|\hat{\mathbf{t}}\|},$$

а также угол между истинным значением поворота \mathbf{R}_{true} и оценкой $\hat{\mathbf{R}}$ как меру ошибки

$$\epsilon_R = \arccos\left(\frac{\operatorname{tr}(\hat{\mathbf{R}}^{-1}\mathbf{R}_{true}) - 1}{2}\right).$$

Результаты показаны на рисунке 1

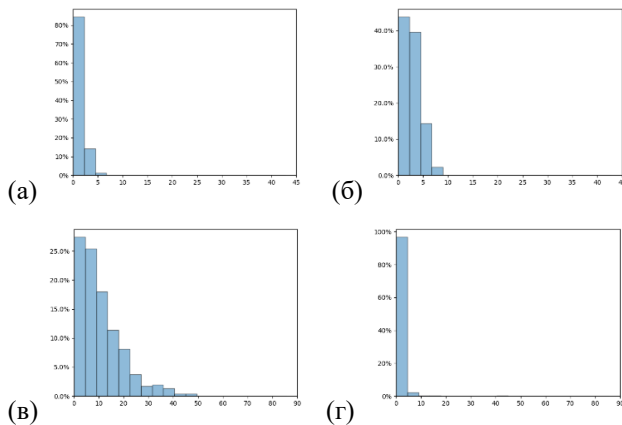


Рис. 1. Распределение ошибок полученных оценок на синтетических данных: (а) поворот, декомпозиция существенной матрицы, (б) поворот, предложенный подход, (в) сдвиг, декомпозиция существенной матрицы, (г) сдвиг, предложенный подход

Легко заметить, что, хотя распределение ошибки вращения несколько хуже для предложенного подхода, оно остается в допустимом диапазоне значений. С другой стороны, при использовании предлагаемого подхода значительно снижается ошибка оценки сдвига. В целом, предлагаемый метод обеспечивает большую надежность.

Также предложенный метод был протестирован на наборе данных MIT Blackbird [5]. Эксперименты проводились на парах изображений из набора данных. Для получения соответствующей пары точек была использована стандартная процедура, включающая обнаружение и сопоставление точек. На рисунке 2 показано распределение ошибок перевода для оценки с использованием декомпозиции существенной матрицы и предлагаемого метода.

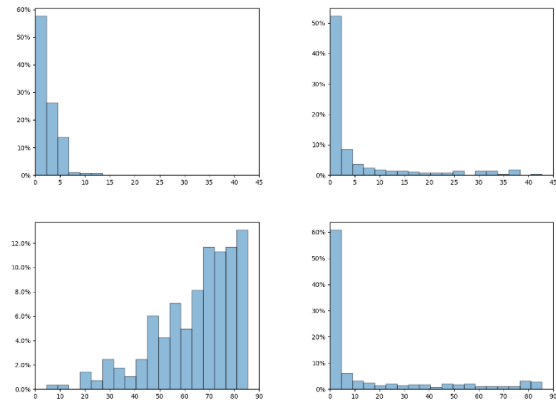


Рис. 2. Распределение ошибок полученных оценок на данных MIT Blackbird: (а) поворот, декомпозиция существенной матрицы, (б) поворот, предложенный подход, (в) сдвиг, декомпозиция существенной матрицы, (г) сдвиг, предложенный подход

Показано, что традиционный подход к оцениванию методом декомпозиции существенной матрицы демонстрирует около 95% неправильных оценок поворота (более 30 градусов). Предлагаемый подход дает приемлемые результаты более чем в 75% случаев.

5. ЗАКЛЮЧЕНИЕ

Предложенный подход к оценке движения камеры показывает гораздо более стабильные результаты для сложных сцен, когда количество соответствующих точек невелико, и большинство этих точек лежат в одной плоскости или в одной и той же части изображения.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках государственного задания по теме 0777-2020-0017 и программы «Приоритет-2030».

ЛИТЕРАТУРА

- [1] He, M. A review of monocular visual odometry / M. He, C. Zhu, Q. Huang, B. Ren, J. Liu // The Visual Computer. – 2019. – Vol. 36(5). – P. 1053-1065. DOI: 10.1007/s00371-019-01714-6.
- [2] Scaramuzza, D. Visual Odometry / D. Scaramuzza, F. Fraundorfer // IEEE Robotics' Automation Magazine. – 2011. – Vol. 18(4). - P. 80-92. DOI: 10.1109/mra.2011.943233.
- [3] Longuet-Higgins, H. A computer algorithm for reconstructing a scene from two projections / H. Longuet-Higgins // Nature. – 1981. – Vol. 293(5828). – P. 133-135. DOI: 10.1038/293133a0.
- [4] Goshin, Y. Estimating Intrinsic Camera Parameters Using the Sum of Cosine Distances / Y. Goshin // Journal of Physics: Conference Series. – 2018. – Vol. 1096. – P. 012092. DOI: 10.1088/1742-6596/1096/1/012092.
- [5] Antonini, A. The Blackbird UAV dataset / A. Antonini, W. Guerra, V. Murali, T. Sayre-McCord, S. Karaman // The International Journal of Robotics Research. – 2020. – Vol. 39(10-11). – P. 1346-1364. DOI: 10.1177/0278364920908331.

Сравнение алгоритмов отбора признаков для задач классификации данных

М.Д. Тисленко

Самарский национальный исследовательский университет
им. академика С.П. Королева
Самара, Россия
makstislenko@gmail.com

А.В. Гайдель

Институт систем обработки изображений - филиал
ФНИЦ «Кристаллография и фотоника» РАН
Самара, Россия
andrey.gaidel@gmail.com

Аннотация—В данной статье рассматриваются различные алгоритмы отбора признаков, производится сравнение качества классификации с использованием алгоритмов отбора и без их использования на различных наборах данных.

Ключевые слова— отбор признаков, алгоритм, классификация.

1. ВВЕДЕНИЕ

Отбор признаков, как один из способов предварительной обработки данных, доказал свою эффективность и действенность при подготовке информации (особенно данных с большим числом признаков) для различных задач интеллектуального анализа данных и машинного обучения. В отбор признаков входит построение более простых и понятных моделей, улучшение качества интеллектуального анализа данных, производительности и подготовка более понятного для восприятия набора признаков [1].

Существует огромное количество различных методов определить лучшее подмножество признаков. Это связано с тем, что данная задача является NP-трудной, гарантировано оптимальное решение может быть найдено только путем полного перебора, который может занимать много времени при большом числе атрибутов [2].

Задачей данного исследования является сравнение различных способов отбора признаков для совершенствования выбора алгоритма для решения задач классификации данных.

2. СРАВНЕНИЕ АЛГОРИТМОВ ОТБОРА ПРИЗНАКОВ

Для сравнения различных способов отбора признаков было взято 4 набора данных [3]-[6]. В качестве классификатора было взято две модели: логистическая регрессия, которая работает быстро, по сравнению с другими моделями и случайный лес, который работает сравнительно точно относительно других моделей, но гораздо медленнее.

В исследовании сравниваются алгоритм SelectKBest с тремя критериями отбора: information gain, Хи-квадрат, F-статистика дисперсионного анализа и алгоритм отбора на основе случайного леса RandomForest. Алгоритм SelectKBest оценивает по отдельности каждый признак по одному из критериев, указанных выше, далее выбирается k лучших признаков. Данный алгоритм считается одним из самых простых и быстрых среди алгоритмов отбора признаков. RandomForest основан на построении так называемых деревьев решений, обучающихся на основе выбранного случайного подмножества признаков, таким

образом, он относится к большой группе ансамблевых методов классификации. После обучения результирующее значение будет браться как среднее от результата работы всех решающих деревьев. Данный метод позволяет производить достаточно точную классификацию, кроме того, можно явно указывать количество признаков, которое необходимо взять для классификации. Таким образом, с помощью данного алгоритма можно не просто классифицировать, но и успешно отбирать признаки. В данном исследовании этот алгоритм используется как в качестве классификатора, так и в качестве алгоритма отбора признаков.

Число выбранных свойств не превышает 10% от изначального количества. Оценка эффективности каждого из алгоритмов проводилась с помощью сравнения взвешенного среднего F-меры по каждому из возможных результирующих значений.

F-мера достигает наилучшего значения при 1 и наихудшего при 0, вычисляется по следующей формуле:

$$F=2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall}) \quad (1)$$

С помощью использования взвешенного среднего при оценке критерия или алгоритма можно принимать во внимание количество элементов, принадлежащих каждому из классов.

$$\bar{x} = \frac{\sum_k w_k x_k}{\sum_k w_k} \quad (2)$$

Где w_k - веса, а x_k - соответствующие этим весам значения. Весом в нашем случае является количество отнесенных к этому классу объектов, так как классов в каждом из наборов данных два, можно переписать формулу в следующем виде:

$$\bar{x} = (w_0 x_0 + w_1 x_1) / (w_0 + w_1) \quad (3)$$

Ниже в таблицах представлены результаты работы каждого алгоритма отбора признаков на наборах данных [3]-[6] и соответствующих классификаторах. На диаграммах по оси ординат представлены значения взвешенного среднего F-меры результатов классификации для каждого из наборов данных при использовании различных методов отбора и без них, по оси абсцисс указаны наборы данных, на которых проводится классификация.

На диаграмме видно, что наиболее качественная классификация с помощью логистической регрессии происходит при отборе признаков с помощью случайного леса, доля верно классифицированных объектов примерно на 2% выше, чем при классификации без отбора

признаков, а на наборе данных The broken machine взвешенное среднее F-меры выше более чем на 5%.

ТАБЛИЦА 1. РЕЗУЛЬТАТЫ РАБОТЫ ПРИ ИСПОЛЬЗОВАНИИ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

	Email spam	Parkinson disease	The broken machine	Financial indicators
Без отбора	0.5751	0.5206	0.7617	0.9562
Хи-квадрат	0.5469	0.5078	0.7931	0.9474
F-статистика ANOVA	0.5837	0.5202	0.7961	0.9403
RandomForest	0.5934	0.5199	0.8146	0.9595
Взаимная информация	0.5	0.5262	0.7816	0.952

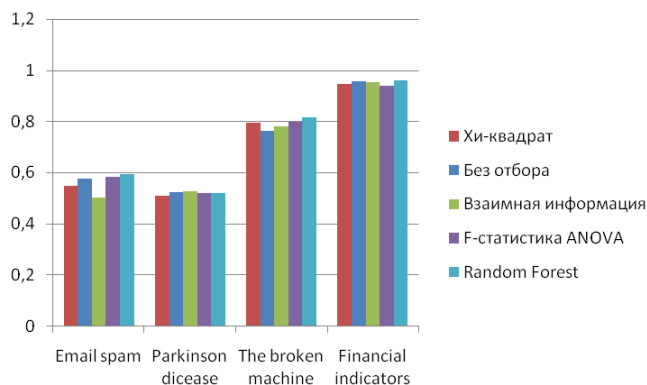


Рис. 1. Диаграмма работы различных методов с использованием логистической регрессии

На остальных наборах данных преимущество в доле правильно классифицированных объектов у RandomForest не такое большое. Видно, что использование критерия взаимной информации на наборе данных Email spam ухудшает качество классификации по сравнению с классификацией без отбора признаков. В целом результаты при использовании отбора и без него практически не отличаются, однако можно отметить, что стабильно качественная классификация также возможна при использовании критериев Хи-квадрат и F-статистики дисперсионного анализа.

На диаграмме видно, что наиболее качественная классификация с помощью случайного леса происходит при отборе признаков с помощью критерия взаимной информации, доля верно классифицированных объектов примерно на 2,5% выше, чем при классификации без отбора признаков.

ТАБЛИЦА 2. РЕЗУЛЬТАТЫ РАБОТЫ ПРИ ИСПОЛЬЗОВАНИИ RANDOM FOREST

	Email spam	Parkinson disease	The broken machine	Financial indicators
Без отбора	0,61295	0,570025	0,830825	0,9614
Хи-квадрат	0,6217	0,5748	0,8704	0,9676
Взаимная информация	0,6304	0,6163	0,8667	0,964
F-статистика ANOVA	0,6012	0,5877	0,8507	0,9675
Random Forest	0,6199	0,5698	0,83	0,9604

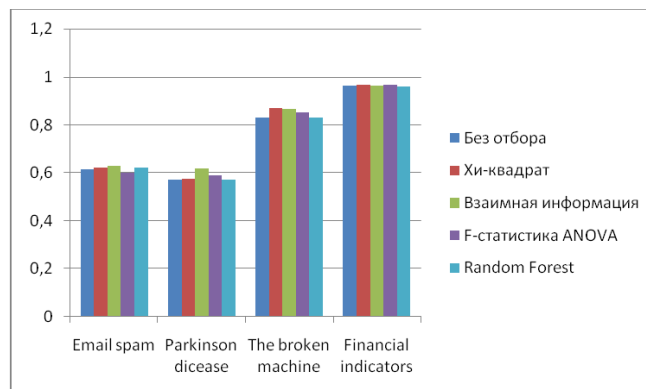


Рис. 2. Диаграмма работы различных методов с использованием классификатора Random Forest

На наборе данных Parkinson Disease этот статистический критерий позволяет улучшить результат классификации на 4,5% по сравнению с классификацией без использования отбора признаков. На остальных наборах данных также есть некоторые улучшения при классификации. Важно отметить, что отбор признаков с помощью RandomForest при использовании классификатора на основе случайного леса показывает результат хуже по сравнению со статистическими критериями, чем при использовании этого же классификатора в сочетании с алгоритмом SelectKBest. В целом результаты при использовании отбора и без него практически не отличаются, однако можно отметить, что на всех наборах данных достаточно качественная классификация происходит при использовании критерия Хи-квадрат и критерия взаимной информации.

3. ЗАКЛЮЧЕНИЕ

В результате работы были использованы различные алгоритмы и критерии отбора признаков для классификации данных. В среднем отбор признаков дает небольшое улучшение классификации. Важно отметить, что использование некоторых критериев на некоторых наборах данных может дать худшее качество классификации, чем было до этого без отбора признаков. На основании результатов из таблиц можно сделать вывод, что наиболее эффективным является отбор признаков с помощью RandomForest, результаты отбора с помощью этого алгоритма почти на всех наборах данных и классификаторах лучше, чем без использования отбора. Кроме того, отбор признаков этим методом дал наибольший прирост в взвешенного среднего F-меры по сравнению с классификацией без отбора признаков. В среднем доля верно классифицированных объектов увеличилась на 1% при использовании этого алгоритма. Также можно отметить, что стабильно высокое качество классификации возможно при использовании критерия Хи-квадрат в алгоритме выбора k наилучших признаков. Остальные критерии, которые можно использовать при выборе k наилучших признаков также в среднем показывают результаты, которые незначительно лучше, чем без отбора, однако при выборе определенных наборов данных и классификаторов качество классификации может быть на несколько процентов хуже, чем без отбора признаков.

Таким образом, можно сделать вывод, что отбор признаков с помощью случайного леса в общем случае при невозможности подробного анализа набора данных

покажет наилучший результат, однако может оказаться так, что классификация происходит точнее на данном наборе данных при выборе k наилучших с помощью какого-либо статистического критерия, однако выбор критерия требует подробного изучения набора данных, что не всегда осуществимо.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке гранта РФФИ, проект № 20-51-05008 Агм_а

ЛИТЕРАТУРА

- [1] Li, J. Feature Selection: A Data Perspective/ J. Li, K. Cheng, S. Wang, F. Morstatter, R. Trevino, J. Tang, H. Liu // ACM Computing Surveys. – 2017. – Vol. 50 – P. 94-139.
- [2] Ходашинский, И.А. Отбор классифицирующих признаков: сравнительный анализ бинарных метаэвристик и популяционного

алгоритма с адаптивной памятью / И.А. Ходашинский, К.С. Сарин // Программирование. – 2019. – Т. 45, № 5. – С. 3-9. DOI: 10.1134/S0132347419050030.

- [3] Email spam classification dataset [Electronic resource]. — Access mode: <https://www.kaggle.com/balaka18/email-spam-classification-dataset-csv> (05.02.2022).
- [4] Parkinson disease speech signal features [Electronic resource]. — Access mode: <https://www.kaggle.com/dipayanbiswas/parkinsons-disease-speech-signal-features> (05.02.2022).
- [5] The broken machine [Electronic resource]. — Access mode: <https://www.kaggle.com/ivanloginov/the-broken-machine> (05.02.2022).
- [6] 200 + financial indicators of US stocks (2014-2018) [Electronic resource]. — Access mode: <https://www.kaggle.com/cnic92/200-financial-indicators-of-us-stocks-20142018> (05.02.2022).

Распределенная реализация метода средних решающих правил с умными выборками для больших задач SVM

М.Ю. Курбаков
Тульский государственный университет
Тула, Россия
muwsik@mail.ru

С.Д. Двоенко
Тульский государственный университет
Тула, Россия
sergedv@yandex.ru

А.В. Копылов
Тульский государственный университет
Тула, Россия
andkopylov@gmail.com

В.В. Сулимова
Тульский государственный университет
Тула, Россия
vsulimova@yandex.ru

Аннотация—Ранее нами был предложен метод средних решающих правил с умными выборками (SS-KMDR) для эффективного решения двухклассовых задач обучения по методу опорных векторов (SVM) в условиях большого числа объектов и было экспериментально показано, что он позволяет получить результаты, близкие по качеству к эталонным за существенно меньшее время по сравнению с традиционными методами. В данной работе мы представляем его высокопроизводительную реализацию, которая позволяет дополнительно существенно повысить скорость вычислений за счет разработки эффективных параллельных алгоритмов и применения современных технологий параллельных и распределенных вычислений.

Ключевые слова— SVM, большая обучающая совокупность, высокопроизводительные вычисления.

1. ВВЕДЕНИЕ

При решении больших задач SVM возникают проблемы высокой вычислительной сложности, нехватки памяти и быстрого произвольного доступа к объектам, осложняющиеся тем, что традиционный формат хранения данных (libsvm) не позволяет вычислить положение объекта с заданным номером.

Практически все существующие методы, детальный обзор которых приведен в [1], направлены на решение только одной из указанных проблем, а также, как правило, имеют и другие недостатки, наиболее важными из которых является отсутствие возможности введения нелинейности и итерационная природа с многочисленными зависимостями по данным.

Предложенный нами ранее метод средних решающих правил с умными выборками (SS-KMDR) [1] лишен указанных недостатков и позволяет получить результаты, близкие по качеству к эталонным за ощутимо меньшее время. В данной работе мы представляем его распределенную версию, обеспечивающую дополнительное повышение производительности.

2. МЕТОД СРЕДНИХ РЕШАЮЩИХ ПРАВИЛ С УМНЫМИ ВЫБОРКАМИ (SS-KMDR)

А. Основная идея

Основная идея заключается в формировании небольших обучающих подвыборок, независимом обучении по каждой из них с последующим

объединением в общее решение исходной задачи. В [1] показано, что и в пространстве признаков, и в пространстве, порожденном потенциальной функцией, объединение может быть осуществлено путем усреднения, но необходимо специфическое понимание усреднения для каждого из этих случаев.

В отличие от традиционных для выборочных методов [2], [3] случайных обучающих подвыборок предлагается использовать специально сформированные (умные) выборки, составленные из объектов, расположенных вблизи разделяющей гиперплоскости, а именно, опорных объектов, полученных при обучении по небольшим случайным подвыборкам.

Б. Оптимизация работы с данными

В [1] предложена специальная схема, основанная на осуществлении предварительной разметки файла с данными для определения расположения в нем объектов. Предварительная разметка в совокупности с применением механизма отображения файла в память, позволяет осуществить быстрый произвольный доступ к любому объекту, не требуя одновременной загрузки в память всех объектов и снимая практическое ограничение на размер обучающей совокупности.

В. Двухуровневая схема параллельных вычислений

Метод SS-KMDR может быть алгоритмически реализован несколькими способами. В данной работе используется версия с фиксированным числом умных выборок. Исходными параметрами в этом случае являются: количество умных выборок, размер одной умной выборки, размер одной случайной подвыборки и полная обучающая совокупность.

Поскольку каждая умная выборка может быть сформирована и обработана независимо от других, то известное заранее количество умных выборок позволяет легко реализовать массивный параллелизм верхнего уровня при помощи процессов. Процессы могут работать на одном или нескольких узлах вычислительной системы. Последнее является более предпочтительным, поскольку позволяет выделить большее количество ресурсов для каждой задачи.

Задача формирования одной умной выборки, в свою очередь, может быть разбита на ряд подзадач формирования случайных подвыборок и обучения по

ним. Однако необходимое количество случайных подвыборок нельзя определить заранее, так как неизвестно сколько опорных объектов будет получено в результате обучения по каждой из них. Это препятствует организации параллельного выполнения данного этапа.

Для решения этой проблемы предлагается смешанная (последовательно-параллельная) схема формирования умной выборки, реализуемая при помощи потоков. Эта схема заключается в том, чтобы сформировать и параллельно решить пул задач обучения на случайных подвыборках, объединить полученные опорные объекты и, если их количество меньше заданного объема умной выборки, сформировать и решить новый пул задач и т. д., пока не будет достигнут размер умной выборки.

3. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

Исследование проводилось с использованием суперкомпьютера НИИ ВЦ МГУ «Ломоносов-2» [4] на наборах данных из репозитория LibSVM (<https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/>) и PiSVM (<https://sourceforge.net/projects/pisvm/files/pisvm-datasets/pisvm-dataset-1.0/>), их основные характеристики приведены в таблице I.

ТАБЛИЦА 2. РЕЗУЛЬТАТЫ СРАВНЕНИЯ SS-KMDR С ДРУГИМИ МЕТОДАМИ ОБУЧЕНИЯ SVM

Метод	Набор данных							
	mnist-784		kddcup99		SUSY		HIGGS	
	время (с)	точность	время (с)	точность	время (с)	точность	время (с)	точность
LibSVM	387	99.15	> 3600	-	> 86 400	-	> 604 800	-
PiSVM	302	99.15	Нехватка ОП		Нехватка ОП		Нехватка ОП	
Бэггинг-1	157	99.47	2241	92.74	> 86 400	-	> 86 400	-
Бэггинг-5	775	99.49	> 3600	-	> 86 400	-	> 86 400	-
SGD	0.633	94.39	16.98	91.96	40.5	77.14	230	62.55
ASGD	0.675	94.47	15.97	91.95	35.2	78.10	144	63.62
SS-KMDR (1/1; 5000)	6.24	98.54	0.358	91.95	5.27	81.12	23.6	65.24
SS-KMDR (8/1; 5000)	58	99.45	3.29	92.35	72	84.32	213	70.23
SS-KMDR (8/8; 5000)	8.06	99.47	0.479	92.33	6.82	84.35	26.7	70.22
SS-KMDR (8/8; 20000)	-	-	0.74	92.57	63.6	86.93	170	75.41
SS-KMDR (16/16; 20000)	-	-	0.89	92.64	66.2	87.27	173	80.05

4. ЗАКЛЮЧЕНИЕ

Эксперименты показывают, что классические LibSVM, его параллельная версия PiSVM и Бэггинг не позволяют получить решение для большинства рассмотренных наборов данных из-за ограничений по времени и памяти. Быстрые методы SGD и ASGD не допускают введение нелинейности, поэтому проигрывают по качеству. Предложенный подход позволяет достаточно быстро найти решение, близкое к эталонному, а его параллельная реализация обеспечивает дополнительное повышение производительности, характеризуется высокой эффективностью использования процессоров и, более того, в ряде случаев позволяет повысить качество решения за счет охвата большего объема данных за меньшее время.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке РФФИ, проекты № 20-07-00055 и № 20-07-00441 с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ имени М.В. Ломоносова [4].

ТАБЛИЦА 1. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ НАБОРОВ ДАННЫХ

Набор данных	Объектов на обучении/ контроле	Число признаков (доля не нулевых)
mnist-784	60 000 / 10 000	784 (61.73%)
kddcup99	4 898 430 / 11 029	122 (10.23%)
SUSY	4 500 000 / 500 000	18 (98.82 %)
HIGGS	10 500 000 / 500 000	28 (92.11 %)

В таблице II приведены время работы и точность распознавания на контрольной выборке для предложенного метода (SS-KMDR), библиотеки LibSVM [5] (являющейся эталоном качества), ее параллельной версии PiSVM [6], а также python scikit-learn реализацией бэггинга [2] с ансамблем из 1 и 5 моделей, а также стохастических методов SGD [7] и ASGD [8]. Результаты для метода SS-KMDR приведены для нескольких наборов значений параметров, указанных в скобках в следующем формате: (<число умных выборок> / <число процессов>; <размер умной выборки>). Размер случайной выборки везде был принят равным 300. Для остальных методов были взяты параметры, принятые по умолчанию. Для mnist-784 для SS-KMDR (*/*; 20000) вычисления не проводились, из-за относительно небольшого числа объектов.

ЛИТЕРАТУРА

- [1] Makarova, A. Mean Decision Rules Method with Smart Sampling for Fast Large-Scale Binary SVM Classification / A. Makarova, M. Kurbakov, V. Sulimova // ICPR. – 2021. – P. 8212-8219.
- [2] Breiman, L. Bagging predictors / L. Breiman // ML. – 1996. – Vol. 24(2). – P. 123-140.
- [3] Chauhan, V.K. Mini-batch block-coordinate based stochastic average adjusted gradient methods to solve big data problems / V.K. Chauhan, D. Kalpana, S. Anuj // Asian Conf. on Machine Learn., PMLR. – 2017. – P. 49-64.
- [4] Voevodin, VI. Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community / VI. Voevodin, A. Antonov, D. Nikitenko, P. Shvets, S. Sobolev, I. Sidorov, K. Stefanov, Vad. Voevodin, S. Zhumatiy // Supercomputing Frontiers and Innovations. – 2019. – Vol. 6(2). – P. 4-11. DOI:10.14529/jsfi190201.
- [5] Chang, C.-C. LIBSVM: a library for support vector machines / C.-C. Chang, C.-J. Lin. – 2001. URL: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [6] Chang, E.Y. Psvm: Parallelizing support vector machines on distributed computers / E.Y. Chang, K. Zhu, H. Wang, H. Bai, J. Li, Z. Qiu, H. Cui / NIPS. – 2007. – Vol. 20.
- [7] Bottou, L. Stochastic Learning / L. Bottou, O. Bousquet, U. von Luxburg // Advanced Lect. on ML, LNAI 3176. – Springer Verlag, Berlin, 2004. – P. 146-168.
- [8] Averaging Stochastic Gradient Descent on Riemannian Manifolds / N. Tripuraneni, N. Flammarion, F. Bach, M. Jourdan // arXiv:1802.0912, 2018.

Анализ социально-экономического развития региона на базе цифровой платформы интегрального мониторинга

П.В. Ситников
ООО «Открытый код»
Самара, Россия
sitnikov@o-code.ru

Е.А. Додонова
Самарский государственный
технический университет
Самара, Россия
dodonova.evg@gmail.com

И.Н. Дубинина
Самарский государственный
технический университет
Самара, Россия
dubinina@o-code.ru

А.В. Ивашенко
Самарский государственный
технический университет
Самара, Россия
anton.ivashenko@gmail.com

О.К. Головнин
Самарский национальный
исследовательский университет имени
академика С.П. Королева
Самара, Россия
golovnin@bk.ru

Н.Ю. Ильясова
Институт систем обработки
изображений - филиал ФНИЦ
«Кристаллография и фотоника» РАН
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия

Аннотация—В статье представлена современная разработка в области повышения эффективности бизнес-аналитики на основе реализации современных технологий семантического и статистического анализа данных открытых источников и государственных информационных систем. Представленная платформа содержит в своем составе инструменты обработки больших данных, компоненты систем поддержки принятия решений и инструментальные средства анализа информации на основе искусственного интеллекта. Практическая реализация платформы позволяет усовершенствовать анализ показателей социального и экономического развития региона.

Ключевые слова— цифровая трансформация, геоинформационная платформа, социальная среда, мониторинг и анализ данных, показатели, поддержка принятия решений.

1. ВВЕДЕНИЕ

Современные тенденции по цифровой трансформации государственного управления предусматривают сбор и актуализацию данных о темпе социально-экономического развития региона. В рамках этой деятельности происходит расчет основных показателей развития, а в качестве исходных данных используется как официальная статистика, так и данные открытых источников информации, таких как новостные ленты и социальные сети.

При этом важно не только обеспечить своевременный и достоверный сбор данных и адекватную их визуализацию, но и реализовать поддержку принятия решений с определением влияющих факторов и выработкой рекомендаций. Данные возможности были реализованы в цифровой платформе интегрального мониторинга, представленной в данной статье.

2. ПРОБЛЕМЫ МОНИТОРИНГА РЕГИОНАЛЬНОГО РАЗВИТИЯ

Интегрированные информационные системы играют ключевую роль в процессе цифровой трансформации за счет создания современной высокоскоростной инфраструктуры хранения, анализа и передачи данных,

собираемых из различных источников в больших объемах [1 - 3]. Для обработки информации обычно применяются технологии искусственного интеллекта и алгоритмы обработки больших данных, что позволяет эффективно анализировать разнородные данные любых размеров в распределенной сетевой среде. Сочетание семантического и статистического анализа данных позволяет реализовать мониторинг ситуации на основе знаний [4, 5].

Среди проблем мониторинга показателей социально-экономического развития региона можно выделить следующие:

- данные официальных источников статистики не отражают реального уровня социально-экономического развития;
- высокая трудоемкость ручного ввода и влияние человеческого фактора;
- недостоверность и рассогласованность данных, получаемых из разных источников;
- отсутствие интеллектуальной обработки данных, прогноза и поддержки принятия решений;
- отсутствие интеграции между ведомственными системами;
- трудность в сравнении и сопоставлении разных показателей.

3. ОПИСАНИЕ РАЗРАБОТАННОЙ ЦИФРОВОЙ ПЛАТФОРМЫ

Цифровая платформа интегрального мониторинга представляет собой геоинформационную систему с настраиваемыми информационными дашбордами, описывающими динамику показателей развития текущей обстановки. Она предназначена для мониторинга результатов деятельности органов исполнительной власти и целевых показателей на основе современных информационных технологий агрегации, обработки и анализа больших данных.

Платформа в своем составе содержит компоненты бизнес-анализа, инструменты обработки больших данных, компоненты систем поддержки принятия

решений и инструментальные средства на основе искусственного интеллекта.

В числе основных задач, решаемых системой, можно выделить следующие:

- автоматизированная загрузка данных из разных источников;
- автоматизированная выгрузка отчетных сведений напрямую в федеральные системы;
- мониторинг показателей социально-экономического развития для формирования целостной и объективной картины;
- прогноз социально-экономического развития;
- выявление закономерностей и корреляционных связей между данными целевых показателей;
- выявление развития кризисных ситуаций по их предпосылкам для оперативного реагирования при их возникновении;
- модуль поддержки принятия решений по улучшению показателей;
- формирование статистических и исторических отчетов на основе полученных данных.

В платформе присутствует возможность графического отображения показателей на карте, а также всех сформированных событий для проведения мероприятий по устранению или профилактике проблемных ситуаций на территории региона. Также анализируется общественное мнение, привязанное к исследуемым показателям. Оно собирается из открытых источников и классифицируется по шкале, отражающей соотношение негативных, нейтральных и положительных сообщений. Данное решение позволяет оценить влияние тех или иных мероприятий, проводимых по федеральным и региональным целевым программам, направленных на изменение социального самочувствия граждан.

Выявление связей между показателями и их визуализация происходит путем работы интеллектуального модуля «Открытый куб». Модуль поддержки принятия решений выдает перечень рекомендаций по достижению необходимых значений целевого показателя на основании оперативной

информации, выявленных взаимосвязей между показателями и данных, полученных в результате прогнозирования. Принятие эффективных управленческих решений обеспечивается применением оперативно функционирующего аналитического аппарата, который позволяет находить причинно-следственные связи в явлениях, прогнозировать ситуации, выработать альтернативные варианты решений, оценивать их и контролировать исполнение.

4. ЗАКЛЮЧЕНИЕ

Цифровая платформа интегрального мониторинга позволяет выявлять скрытые закономерности и определять предпосылки к их возникновению. Наглядная аналитика, отражающая реальное состояние дел, дает реальную картину происходящего в разных формах представления и позволяет сформировать адекватные рекомендации, направленные на предупреждение рисков возникновения инцидентов.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке РФФИ (грант № 19-29-01135), в рамках выполнения работ по государственному заданию ФНИЦ «Кристаллография и фотоника» РАН.

ЛИТЕРАТУРА

- [1] Digital Russia. New Reality // Digital McKinsey. – 2017. – 133 p [Электронный ресурс]. – Режим доступа: <https://www.mckinsey.com/ru/our-work/mckinsey-digital>.
- [2] Седойкина, А.А. Аналитическая обработка данных. Обзор BI-платформ / А.А. Седойкина // Контентус. – 2019. – № S11. – С. 96-102.
- [3] Srivastava, G.S. A review of the state of the art in business intelligence software / G.S. Srivastava, M. Venkataraman, R.V. Kavitha, N. Parthiban // Enterprise Information Systems. – 2021. – P. 128.
- [4] Surmin, O.L. Industrial application of Big Data services in digital economy / O.L. Surmin, P.V. Sitnikov, A.A. Khorina, A.V. Ivaschenko, A.A. Stolbova, N.Yu. Ilyasova // CEUR Workshop Proceedings. – 2019. – Vol. 2416. – P. 409-416.
- [5] Sumin, O. Data exchange platform for digital economy applications / O. Sumin, P. Sitnikov, A. Khorina, A. Ivaschenko, A. Stolbova, N. Ilyasova // CEUR Workshop Proceedings. – 2020. – Vol. 2667. – P. 243-247.

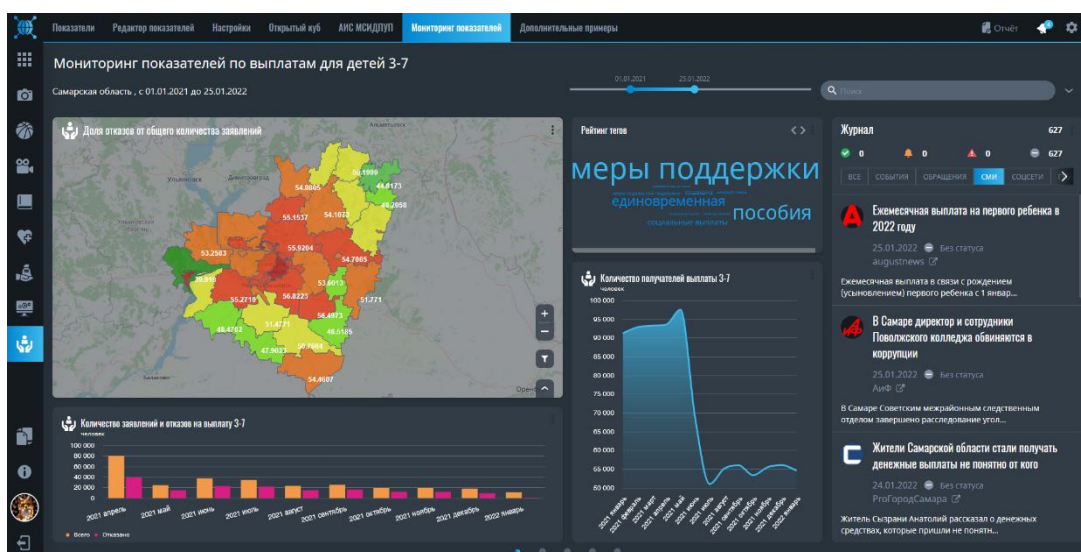


Рис. 1. Цифровая платформа интегрального мониторинга

Применение технологий удаленных вычислений для исследования физико-химических процессов

Э.Н. Мифтахов
Башкирский государственный
университет
Уфа, Россия
promif@mail.ru

С.А. Мустафина
Башкирский государственный
университет
Уфа, Россия
mustafina_sa@mail.ru

Т.А. Михайлова
Башкирский государственный
университет
Уфа, Россия
t.a.mikhailova@yandex.ru

А.Х. Даминов
Башкирский государственный
университет
Уфа, Россия
azatdaminov@toltekplus.ru

Аннотация—Оценка технологических свойств продуктов физико-химической отрасли с применением средств математического моделирования осложняется отсутствием достаточной ресурсной и вычислительной базы, поскольку большой интерес вызывает необходимость осуществления серии расчетных задач для различных исходных данных. Эффективной в этом случае является организация системы удаленных вычислений средствами высокопроизводительного сервера. Актуальной задачей в этих условиях является разработка унифицированных методов и алгоритмов для организации системы удаленных вычислений с целью моделирования процессов физико-химической отрасли. Кроме того, предлагаемый подход способствует формированию единой базы лабораторных/вычислительных экспериментов, позволяющей ставить и решать перспективные задачи применения нейросетевых технологий в задачах химической кинетики.

Ключевые слова— облачные технологии, моделирование, физическая химия.

1. ВВЕДЕНИЕ

Разработка высокоэффективных технологий в химико-технологической отрасли предполагает создание новых методов, позволяющих не только проводить эмпирическую оценку качества получаемого продукта, но и решать задачи проектирования энергосберегающих химико-технологических процессов и систем с целью получения продуктов производства с заданными свойствами. Большую актуальность в таком случае приобретают моделирующие программы и алгоритмы, которые позволяют оценивать технологические свойства получаемой продукции.

Сложность организации такого подхода обусловлена проблемой математического описания процесса с применением фундаментальных математических моделей и отсутствием достаточной ресурсной и вычислительной базы предприятия. Решение прямой задачи исследования с целью прогнозирования свойств продукта для различных начальных условий требует огромных вычислительных ресурсов. Очевидно, что постановка задач поиска оптимальных режимов ведения производства потребует многократного решения прямой задачи и, как следствие, приводит к невозможности вычисления подобных задач в разумные сроки стандартными средствами организации вычислений.

Ранее авторами работы уже были сделаны шаги по разработке локальных программных продуктов, позволяющих проводить частные исследования физико-химических процессов. На ряд из них имеются свидетельства о регистрации в федеральной службе по интеллектуальной собственности [1,2]. Однако их выполнение осуществляется средствами локальной системы ведения расчетов, без привлечения средств организации системы удаленных вычислений. Целью данного исследования является разработка унифицированных методов и алгоритмов для организации системы удаленных вычислений моделирования физико-химических процессов, которая позволит организовать одновременное выполнение нескольких расчетных задач с гибкой системой распределения вычислительных ресурсов.

2. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ ПРОЦЕССА

При математическом описании сложных физико-химических процессов будем руководствоваться методологией, в соответствии с которой на начальном этапе описывается кинетический механизм ведения реакций, который ложится в основу кинетического и статистического подходов к исследованию процессов.

Кинетический подход (метод дифференциальных уравнений) [3,4], позволяет с применением численных методов решения за короткий вычислительный период получать зависимость усредненных молекулярных характеристик и расхода основных реагентов от времени. Основная масса предлагаемых модельных описаний выполнена с применением кинетического подхода. Кинетический подход к исследованию процессов успешно работает с применением как метода статистических моментов, так и метода производящих функций. Оба способа ориентированы на упрощение вида получаемой системы дифференциальных уравнений и приведение ее размерности к конечному виду.

Статистический подход (метод Монте-Карло) [5] работает по иному принципу и требует обработки и хранения всех макромолекул, участвующих в реакциях. Расчет занимает гораздо больший период, однако позволяет проводить детальное исследование свойств макромолекул. В рамках данного исследования подход его основан на методе, предложенном американским физиком Gillespie [6] для моделирования колебательных

реакций и адаптированном под процессы, описываемых с помощью серии последовательных химических реакций. Алгоритм метода может быть представлен в виде последовательности этапов, формирующих совершение определенного события. В качестве события будем рассматривать осуществление одной из реакций кинетической схемы, а вероятность ее осуществления в данный момент времени будем каждый раз пересчитывать.

3. КОНЦЕПТУАЛЬНАЯ СХЕМА СИСТЕМЫ ОРГАНИЗАЦИИ УДАЛЕННЫХ ВЫЧИСЛЕНИЙ

Система организации сетевого взаимодействия [7] информационной системы для организации удаленных вычислений представлена тремя уровнями организации архитектуры.

Первый слой – клиентское приложение, которое должно выполнять две функции: позволять в удобном виде задавать технологические параметры ведения процесса физико-химического производства и предоставлять доступ к результатам вычисления, включая графическое и табличное представление. Для удобства пользователей, легкого доступа и возможных проблем совместимости представлена веб-интерфейсной оболочкой, написанной с помощью UI-фреймворка Blazor C# [8].

Второй слой – система организации хранения всей информации в базе данных. Основное наполнение формируют результаты вычислений, а также справочная информация, включая кинетические параметры, определяющие скорости химических реакций.

Третий слой (основой) – система организации вычислений, ответственная за выполнение необходимых клиенту расчетов. Этот слой отвечает за все вычисления и должен быть максимально подготовлен к выполнению ресурсоемких задач. Данный слой представлен двумя уровнями организации работы. Первый уровень – управляющая программа, которая взаимодействует с базой данных и в фоновом режиме периодически обращается к ее записям с целью проверки наличия задач на выполнение. Второй уровень содержит в своем ассортименте набор готовых bin файлов, направленных на выполнение определенной задачи из определяемой методологии ведения расчетов. Движение основных потоков данных между интерфейсной частью и системой организации вычислений в виде DFD-диаграммы изображено на рис. 1.

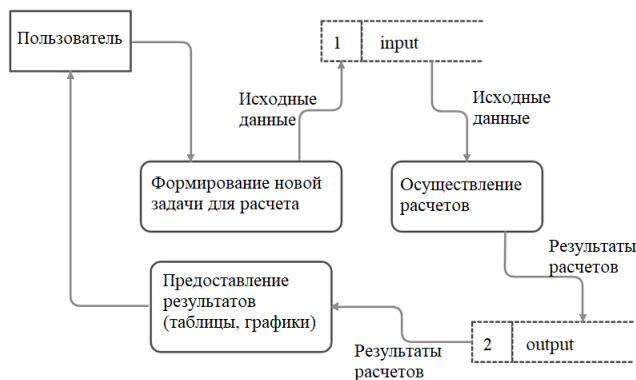


Рис. 1. DFD-диаграмма потоков данных между интерфейсной частью и системой организации вычислений

Возникающая задача визуализации получаемых результатов решена с применением графической библиотеки PlotLY JS, позволяющей производить построение графиков прямо в браузере клиента и не нагружать канал передачи данных. В качестве системы управления базой данных применяется СУБД MongoDB [9], поскольку она достаточно легко масштабируется и поддерживает функцию распределения нескольких таблиц единой базы данных между несколькими серверами.

4. ЗАКЛЮЧЕНИЕ

Предлагаемый подход и разработанная система организации удаленных вычислений позволяет более эффективно использовать вычислительные ресурсы и значительно сократить время выполнения расчетов [10]. Физическое разделение слоя вычислительной логики от слоя хранения данных способствует формированию единой базы лабораторных/вычислительных экспериментов, позволяющей ставить и решать задачи применения нейросетевых технологий в задачах химической кинетики.

БЛАГОДАРНОСТИ

Исследование выполнено в рамках государственного задания Министерства науки и высшего образования Российской Федерации (код научной темы FZWU-2020-0027).

ЛИТЕРАТУРА

- [1] Пат. РФ 2020610226. Решение прямой задачи непрерывного процесса полимеризации изопрена в присутствии микрогетерогенных каталитических систем (опубл. 2020).
- [2] Пат. РФ 2021618232. Решение обратной задачи формирования молекулярно-массового распределения методом регуляризации А.Н. Тихонова (опубл. 2021).
- [3] Подвальный, С.Л. Моделирование промышленных процессов полимеризации / С.Л. Подвальный. – М.: Химия, 1979. – 256 с.
- [4] Мифтахов, Э.Н. Исследование кинетики процесса полимеризации изопрена в присутствии неодимсодержащих каталитических систем, модифицированных в турбулентных потоках / Э.Н. Мифтахов, И.Ш. Насыров, С.А. Мустафина, В.П. Захаров // Журнал прикладной химии. – 2021. – Т. 94, № 1. – С. 81-87. DOI: 10.31857/S0044461821010114.
- [5] Мустафина, С.А. Исследование молекулярных характеристик продукта полимеризации изопрена на неодимсодержащей каталитической системе на основе моделирования методом Монте-Карло / С.А. Мустафина, Т.А. Михайлова, Э.Н. Мифтахов, В.А. Михайлов // Вестник Тверского государственного университета. Серия: Химия. – 2020. – Т. 42, № 4. – С. 138-148.
- [6] Gillespie, D. Exact Stochastic Simulation of Coupled Chemical Reactions / D. Gillespie // The Journal of Physical Chemistry. – 1977. – Vol. 81(25). – P.2340-2361. DOI:10.1021/j100540a008.
- [7] Thomas, E. Cloud Computing Concepts, Technology & Architecture / E. Thomas, Z. Mahmood, R. Puttini. – 2013. – 528 p.
- [8] Roth, D. Blazor for ASP.NET web forms developers (Microsoft Corporation) / D. Roth, J. Fritz, T. Southwick, S. Addie, S. Smith. – Redmond, Washington, 2021. – 98 p.
- [9] Daly, D. Performance Engineering and Database Development at MongoDB / D. Daly // Companion of the ACM/SPEC International Conference on Performance Engineering. – 2021. – 129 p. DOI:10.1145/3447545.3451199.
- [10] Miftakhov, E. Developing methods and algorithms for cloud computing management systems in industrial polymer synthesis processes / E. Miftakhov, S. Mustafina, A. Akimov, O. Larin, A. Gorlov // Emerging Science Journal. – 2021. – Vol. 5(6). – P. 964-972. DOI:10.28991/esj-2021-01324.

Анализ локальных корреляционных характеристик биоэлектрических сигналов человека при выполнении когнитивных задач

В.А. Юнусов
Казанский (Приволжский) федеральный университет
Казань, Россия
valentin.yunusov@gmail.com

С.А. Демин
Казанский (Приволжский) федеральный университет
Казань, Россия
serge_demin@mail.ru

Аннотация—В данной работе в рамках формализма функций памяти проводится автокорреляционный анализ локальных отрезков сигналов электроэнцефалограмм для двух групп людей. Первая группа состояла из 8 музыкантов, профессионально занимающихся музыкой не менее 5 лет, вторая группа была представлена 11 людьми без музыкального образования. Испытуемые выполняли задачи восприятия фрагмента музыкального произведения и восприятия текста. После определения значений автокорреляционных параметров на всех временных отрезках, дополнительно применяется квантильный регрессионный анализ. В результате были установлены области коры головного мозга, в активности которых проявляются наибольшие различия для двух групп людей. Показано, что применение методов статистического анализа позволяет обнаружить различие в когнитивных способностях людей при выполнении различных творческих задач.

Ключевые слова— науки о данных, анализ временных сигналов, локальные параметры, электроэнцефалограммы, автокорреляции, когнитивные задачи.

1. ВВЕДЕНИЕ

В настоящее время в науках о сложности появляются новые возможности для изучения эволюции сложных систем живой и неживой природы благодаря технологиям интеллектуального анализа экспериментальных данных. Этому способствует бурное развитие вычислительной техники и регистрирующего оборудования. В научных работах природа сложных систем раскрывается на примере их уникальных свойств: эмерджентности, самоорганизации, адаптивности, нелинейности. Особое внимание уделяется коллективным эффектам, памяти и корреляциям, которые возникают вследствие взаимодействия между компонентами сложной системы. Коллективные явления оказывают существенное влияние на эволюцию сложных систем: от физики ускорителей (ограничения интенсивности пучка) и физики конденсированных сред (коллективные эффекты) до коллективного поведения в живой природе [1].

Головной мозг человека состоит из десятков миллиардов нейронов, сообщающихся посредством сотни триллионов синапсов. Это сложная система, в которой непрерывно формируются и изменяются корреляционные связи между отдельными частями. Опыт человека и его долгосрочное обучение приводят к изменениям в биоэлектрической активности коры головного мозга, что может быть установлено при проведении статистического анализа записей биомедицинских данных [2, 3]. На данный момент методы количественной и качественной оценки характера

таких изменений интенсивно развиваются благодаря значительному прогрессу теоретических, программных и инструментальных средств исследования. Для решения подобных задач в современных науках о данных широко применяются различные методы статистического анализа сигналов мозговой активности: анализ корреляционных и спектральных характеристик, машинное обучение и нейронные сети [4, 5]. Полученные результаты представляют интерес для наук о сложности, биофизики, нейрофизиологии и когнитивной психологии.

Анализ временных сигналов численными методами позволяет с большей точностью изучить особенности динамических взаимосвязей, которые выстраиваются в ходе эволюции сложных систем.

2. ИССЛЕДОВАНИЕ ЛОКАЛЬНЫХ АВТОКОРРЕЛЯЦИОННЫХ ОСОБЕННОСТЕЙ МОЗГОВОЙ АКТИВНОСТИ

В данной работе разрабатывается оригинальный метод определения отличительных параметров сигналов электроэнцефалограмм (ЭЭГ) людей с различным уровнем развития когнитивных способностей на основе формализма функций памяти [6, 7] и квантильной регрессии – метода, получившего широкое распространение в статистическом анализе для предсказания и исследования временных рядов [8, 9].

В ходе исследования выполнялся анализ корреляционных характеристик записей ЭЭГ для двух групп людей: 8 музыкантов, профессионально занимающихся музыкой не менее 5 лет, и 11 испытуемых без музыкального образования. Записи ЭЭГ фиксировались электродами, размещенными согласно международной системе «10-20%» (Рис. 1), в течение 90 секунд, в состоянии покоя человека с закрытыми глазами. Испытуемые во время проведения записи выполняли когнитивные задачи восприятия фрагмента музыкального произведения и прослушивания текста нейтрального содержания, зачитываемого вслух [10].

На подготовительном этапе записи ЭЭГ сигналов были разделены на временные «окна». Затем для каждого временного окна вычислялся коэффициент автокорреляции. Завершающий этап состоял в количественной оценке зависимости полученных значений от времени методом квантильной регрессии.

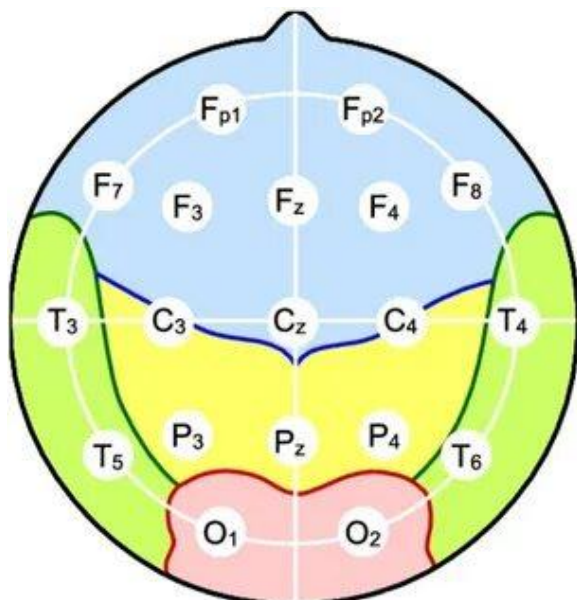


Рис. 1. Расположение электродов согласно международной системе «10-20%» при регистрации электроэнцефалограмм

В ходе регрессионного анализа были установлены электроды, для которых проявлялись наибольшие различия в медианных и квантильных значениях для двух групп испытуемых при выполнении когнитивных задач. В случае задания восприятия музыки наибольшие различия проявлялись в сигналах, регистрируемых электродами T_5 , P_3 , Fp_1 и Fp_2 . В данном случае мы наблюдаем существенные различия в разных частях коры головного мозга человека. В задаче восприятия текста нейтрального содержания наибольшие отличия были обнаружены в областях коры головного мозга, сигналы которых регистрировались электродами T_4 , T_6 , Fp_1 и T_3 , т. е. в наиболее существенной степени были задействованы височная и лобная области.

Следует отметить, что представленные результаты представляют лишь первый шаг в оценке корреляционных взаимосвязей, которые выстраиваются в тех или иных областях коры головного мозга человека в процессе творческой активности. Дальнейшие исследования будут связаны с определением оптимальной длины временных окон [11], вычислением коэффициента кросскорреляции между сигналами, регистрируемыми электродами из разных областей коры головного мозга. Оптимальная длина временного окна позволяет достичь необходимой «чувствительности» рассчитываемых локальных параметров. При малой длине временного окна будет утеряна полезная информация, извлекаемая из сигнала, при большой длине чувствительность будет утеряна за счет «шумовых» эффектов и накоплении погрешностей.

3. ЗАКЛЮЧЕНИЕ

В данной работе в рамках формализма функций памяти и квантильного регрессионного анализа было проведено определение отличительных критериев записей ЭЭГ для испытуемых с профессиональной подготовкой в сфере музыкального искусства и без музыкального образования. Для записей сигналов двух групп испытуемых, разбитых на локальные «окна», был

рассчитан коэффициент автокорреляции, после чего применена квантильная регрессия.

В ходе анализа были установлены области коры головного мозга, для которых различие в динамике биоэлектрической активности фиксировалось в наиболее значительной степени.

Представленные результаты анализа в рамках формализма функций памяти и квантильной регрессии после дополнительной верификации могут быть использованы для разработки интеллектуальной оценки развития когнитивных способностей человека. Для формирования обучающих выборок в этом случае потребуется обработка экспериментальных данных, фиксируемых как для испытуемых с разным уровнем творческих способностей, так и для волонтеров из разных возрастных групп.

БЛАГОДАРНОСТИ

Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»).

ЛИТЕРАТУРА

- [1] Fiegna, F. Evolution of an obligate social cheater to a superior cooperator / F. Fiegna, Y.-T.N. Yu, S.V. Kadam, G.J. Velice / *Nature*. – 2006. – Vol. 441. – P. 310-314.
- [2] Paulo, B. Efficiency of attentional networks in musicians and nonmusicians / B. Paulo, M. David // *Heliyon*. – 2019. – Vol. 5(3). – P. 1-17.
- [3] Habibi, A. Cortical Activity during Perception of Musical Rhythm; Comparing Musicians and Non-musicians / A. Habibi, V. Wirantana, A. Starr // *Psychomusicology*. – 2014. – Vol. 24(2). – P. 125-135.
- [4] Cheung, M. Music training is associated with cortical synchronization reflected in EEG coherence during verbal memory encoding / M. Cheung, A.S. Chan, Y. Liu, D. Law, C.W.Y. Wong // *PLoS One*. – 2017. – Vol. 12(3). – P. 0174906.
- [5] Liang, S. Classification of EEG signals from musicians and nonmusicians by neural networks / S.-F. Liang, T.-H. Hsieh, W.-H. Chen, K.-J. Lin // *IEEE 9th World Congress on Intelligent Control and Automation*. – 2011. – P. 865-869.
- [6] Wan, C. Direct quantile regression for nonparametric probabilistic forecasting of wind power generation / C. Wan, J. Lin, J. Wang, Y. Song, Z. Y. Dong // *IEEE Transactions on Power Systems*. – 2017. – Vol. 32(4). – P. 2767-2778.
- [7] Wu, D. Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR) / D. Wu, V.J. Lawhern, S. Gordon, B.J. Lance, C.-T. Lin // *IEEE Transactions on Fuzzy Systems*. – 2017. – Vol. 25 – P. 1522-1535.
- [8] Demin, S.A. Statistical quantifiers of memory for an analysis of human brain and neuro-system diseases / S.A. Demin, R.M. Yulmetyev, O.Yu. Panishev, P. Hänggi // *Physica A*. – 2008. – Vol. 387(8-9). – P. 2100-2110.
- [9] Panishev, O.Y. Cross-correlation markers in stochastic dynamics of complex systems / O.Y. Panishev, S.A. Demin, J. Bhattacharya // *Physica A*. – 2010. – Vol. 389. – P. 4958-4969.
- [10] Bhattacharya, J. Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise / J. Bhattacharya, H. Petsche // *Signal Processing*. – 2005. – Vol. 85(11). – P. 2161-2177.
- [11] Yulmetyev, R.M. Non-Markov stochastic dynamics of real epidemic process of respiratory infections / R.M. Yulmetyev, N.A. Emelyanova, S.A. Demin, F.M. Gafarov, P. Hänggi, D.G. Yulmetyeva // *Physica A*. – 2004. – Vol. 331(1-2). – P. 300-318.

Выбор стратегии поведения участника электронных торгов в сфере государственных закупок с помощью прогнозирования

Ю.В. Хицкова
Воронежский государственный университет
Воронеж, Россия
prosvetovau@list.ru

И.Ф. Астахова
Воронежский государственный университет
Воронеж, Россия
astachova@list.ru

К.А. Маковий
Воронежский государственный технический университет
Воронеж, Россия
makkatya@mail.ru

О.Е. Ефимова
Воронежский государственный технический университет
Воронеж, Россия
efimova_oe@mail.ru

Аннотация—В статье предложен подход к выбору стратегии участника электронных торгов в сфере государственных закупок, основанный на прогнозировании коэффициента снижения цены контракта с помощью искусственной нейронной сети. Сформулированы стратегии участия в торгах в зависимости от соотношения прогнозируемой цены контракта и себестоимости товаров, работ и услуг

Ключевые слова— стратегия участника электронных торгов, прогнозирование с помощью ИНС, модель NARX.

1. ВВЕДЕНИЕ

Широкое распространение электронных торгов в РФ в сфере государственных закупок вызывает необходимость в исследовании того как накопленные данные, находящиеся в открытом доступе, могут помочь в выборе возможных стратегий поведения конечных участников торговых площадок [1,2]

Несмотря на то, что существует большое количество исследований, посвященных дизайну самих торгов с точки зрения организатора, наблюдается недостаток информации о том, какие существуют стратегии поведения участников электронных торгов, каким образом участник может оценить заказчика с точки зрения перспективы участия в торгах по заявке [3,4].

2. ПРИМЕНЕНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ВЫБОРА СТРАТЕГИИ ПОВЕДЕНИЯ УЧАСТНИКОВ ЭЛЕКТРОННЫХ ТОРГОВ

Нами предлагается выбор стратегии для участника электронных торгов на основе прогнозирования значения временного ряда «коэффициента снижения цены контракта» по данному заказчику. Для прогнозирования значения временного ряда предлагается использовать нейронную сеть глубокого обучения NARX, которая реализована в пакете DeepLearningToolbox системы компьютерной математики Matlab..

А. Получение данных

Одной из задач, которые необходимо решить в первую очередь при проведении любого анализа, является задача получения данных. Сведения о заявках и участниках электронных торгов по государственным закупкам являются публичными. Данные хранятся на FTP-сервере официального сайта единой информационной системы в

сфере закупок в информационно-телекоммуникационной сети Интернет (далее – Официальный сайт ЕИС) и доступны для скачивания в виде заархивированных zip-архивов, внутри которых находятся файлы в формате XML (eXtensible Markup Language — расширяемый язык разметки). Особенностью файлов XML является использование тэгов для определения структуры данных, благодаря чему этот формат очень популярен как универсальный формат передачи данных между разными информационными системами.

Б. Прогнозирование коэффициента снижения цены контракта с помощью нейронной сети

В «голландском» аукционе первоначальная цена, предложенная заказчиком, уменьшается в ходе торгов и целесообразность участия в торгах для организации – участника торгов определяется конечной ценой контракта.

В данной работе мы рассматриваем возможность прогнозирования временного ряда коэффициента снижения цены контракта K_{red} по конкретному заказчику. Временной ряд получен в результате выборки из базы данных значений начальной максимальной цены контракта (НМЦ) P_{max} и цены победителя P_{win} . База данных формируется в результате работы синтаксического анализатора, выполняющего поиск и загрузку данных, доступных на FTP-портале Официального сайта ЕИС. Значения коэффициента снижения цены контракта рассчитывается по формуле:

$$K_{red} = \frac{P_{max} - P_{win}}{P_{max}} \quad (1)$$

Введем обозначение временного ряда $y_1, \dots, y_t, \dots, y_t \in R$, элементы которого рассчитываются по формуле (1). Это вещественный скалярный ряд (рис.1). Выбор модели NARX обусловлен возможностью использовать для увеличения точности прогнозирования экзогенный параметр [5,6], в качестве которого предлагается первоначальная цена закупки или НМЦ, то есть стоимость заявки влияет на коэффициент снижения цены контракта.

Нелинейная авторегрессия с внешним вводом - это модифицированная нелинейная модель авторегрессии, включающая другой соответствующий временной ряд в

качестве дополнительного входа в модель прогнозирования. Модель может быть записана в следующем виде:

$$y_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-d+1}, y_t, y_{t-1}, \dots, y_{t-d+1}) \quad (2)$$

или может быть записана в векторной форме:

$$y_{t+1} = f(x_t, y_t) \quad (3)$$



Рис. 1. Временной ряд коэффициента снижения цены контракта

Здесь x_t – это внешний вход прогнозной модели с таким же количеством временных точек как и y_t , то есть два известных временных ряда используются как независимые входные данные для скрытого слоя в соответствии с тем же числом задержек t . Нелинейное отображение $f(\cdot)$ обычно неизвестно и может быть аппроксимировано через архитектуру модели динамической нейронной сети NARX. Нейронная сеть NARX может быть выражена как $y_t = \sum_i c_i \psi(\sum_{j=1}^d (a_{ji} x_{t-j} + b_{ji} y_{t-j}))$ (4)

где ψ – активационная функция скрытого ряда; a_{ji} и b_{ji} – входные данные для весов скрытого слоя в скрытом нейроне j ; и c_i является скрытым для веса выходного слоя i , d – число входных узлов весов.

Полученные результаты показывают достаточно хорошее совпадение прогнозных данных со значениями рассматриваемого временного ряда.

В. Понятие стратегии и влияние прогнозируемого параметра на выбор стратегии

Разработанная нейронная сеть позволяет прогнозировать коэффициент K_{red} и по нему рассчитать ожидаемую цену товаров, работ и услуг (ТРУ) по заявке. Далее компания сможет принимать решение о дальнейшем участии или отказе о участия в аукционе.

В случае если прогнозируемый коэффициент снижения цены будет слишком велик, прогнозируемая конечная цена окажется ниже себестоимости ТРУ, решение компании будет зависеть от других факторов. Первым фактором является необходимость именно этого заказа организации. Следующим важным фактором

является возможность продать по демпинговой цене. Даже если такая необходимость есть, организация должна обладать высоким коэффициентом финансовой устойчивости, чтобы демпинговать. Третьим фактором может являться конкурентная борьба, насколько организация потеснит конкурентов на рынке благодаря данному заказу. С учетом перечисленных факторов, основываясь на прогнозной цене, компания будет решать проблему целесообразности участия в торгах. Это *пессимистичный* стратегический сценарий. В большинстве случаев целесообразно отказаться от участия в торгах данного заказчика.

Случай, если коэффициент снижения цены покрывает себестоимость организации и принесет некоторую незначительную долю прибыли, можно назвать *реалистичным*. Здесь компания может отказаться от участия в торгах, если прогноз коэффициента снижения цены окажется меньше у других заказчиков. В противном случае ей целесообразно участвовать в аукционе.

Оптимистичный стратегический сценарий предполагает незначительное снижение первоначальной цены и соответственно самый благоприятный прогноз для участия в торгах. Заказчики, прогнозируемый коэффициент падения цены которых является минимальным – наиболее привлекательны, с точки зрения участия в аукционе.

3. ЗАКЛЮЧЕНИЕ

В данной работе показаны преимущества нейросетевого подхода к прогнозированию параметра закупки, на основании которого можно сформулировать различные стратегии для участников электронных торгов, а также выделены факторы, влияющие на их выбор.

ЛИТЕРАТУРА

- [1] Савватеев, А.В. Теория и практика аукционов / А.В. Савватеев, А. Филатов // Вестник Воронежского государственного университета. Серия: Экономика и управление. – 2018. – № 3. – С. 119-131.
- [2] Дук, Г.В. Обзор и применение теории аукционов в мультиагентных системах / Г.В. Дук, А.Н. Быков, С.А. Чернышев // Международный журнал информационных технологий и энергоэффективности. – 2021. – Т. 6, № 1(19). – С. 42-49.
- [3] Rahimiyan, M. Supplier's optimal bidding strategy in electricity pay-as-bid auction: Comparison of the Q-learning and a model-based approach / M. Rahimiyan, H.R. Mashhadi // Electric Power Systems Research. – 2008. – Т. 78, № 1. – С. 165-175.
- [4] Puro, L. Bidding strategies for real-life small loan auctions / L. Puro // Decision Support Systems. – 2011. – Т. 51, № 1. – С. 31-41.
- [5] Astachova, I.F. Possibilities for predicting the state of usability web resources / I.F. Astachova, K.A. Makovy, Yu.V. Khitskova // Journal of Physics: Conference Series: Current Problems. – Voronezh, 2021. – P. 012029. DOI 10.1088/1742-6596/1902/1/012029.
- [6] Metelkin, Y. Predicting Load in Datacenters Using NARX Model / Y. Metelkin, Y. Khitskova, K. Makoviy // International Conference on Information Technology and Nanotechnology (ITNT) – 2021 – P. 1-5. DOI: 10.1109/ITNT52450.2021.9649034.

Двухкаскадные адаптивные подавители помех с управляемыми интервалами адаптации

В.А. Засов

Самарский государственный университет путей сообщения
Самара, Россия
vzasov@mail.ru

П.А. Мельников

Самарский государственный университет путей сообщения
Самара, Россия
gleavero@gmail.com

Аннотация—Предложен двухкаскадный адаптивный подавитель помех в прерывистых и импульсных сигналах. Особенностью устройства является предварительная декорреляция и использование пауз входных сигналов, позволяющие подавлять коррелированные помехи, уровень мощности которых значительно превышает уровень мощности полезных сигналов.

Ключевые слова— подавление, помехи, прерывистый, сигнал, пауза, адаптация, двухкаскадный, декорреляция.

1. ВВЕДЕНИЕ

Классические адаптивные фильтры [1] не эффективны для подавления коррелированных с полезными сигналами помех. Предложенные в [2, 3, 4] адаптивные подавители помех (АПП) с адаптацией только в паузах сигналов позволяют подавлять коррелированные с полезными сигналами помехи. Применение таких АПП [2, 4] для прерывистых сигналов ограничивается уровнем мощности помех, который должен быть меньше уровня мощности полезных сигналов. Если же уровень мощности помех превышает уровень мощности сигналов, полезные сигналы должны быть периодическими [3, 4], что ограничивает область применения этого типа АПП

В работе предлагается алгоритм и реализующий его двухкаскадный АПП в прерывистых и импульсных сигналах, позволяющий подавлять коррелированные с полезными сигналами помехи, уровень мощности которых может значительно превышать уровень мощности сигналов.

2. ДВУХКАСКАДНЫЕ АДАПТИВНЫЕ ПОДАВИТЕЛИ ПОМЕХ В ПРЕРЫВИСТЫХ И ИМПУЛЬСНЫХ СИГНАЛАХ

Предлагаемый в работе АПП отличается от известных [2, 3, 4] тем, что алгоритм подавления помех двухэтапный и состоит из двух ступеней подавления.

На первой ступени производится предварительное подавление коррелированных помех, которое позволяет обнаружить и вычислить паузы в полезном сигнале. На второй (завершающей) ступени производится полное подавление помех с использованием вычисленных на первой ступени пауз [2, 4].

Структурная схема процесса формирования в объекте сигналов, поступающих на входы АПП, приведена на рис. 1, где $s_1(k), s_2(k), s_3(k)$ – источники соответственно полезного сигнала, коррелированной и некоррелированной с полезным сигналом помехи, а $x_1(k)$ и $x_2(k)$ – аддитивные смеси полезного сигнала и помех, поступающее на входы АПП.

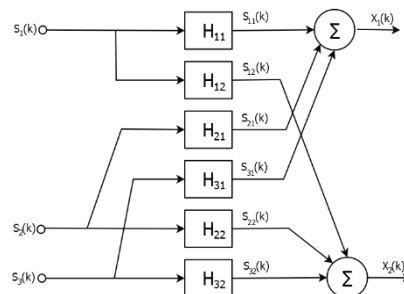


Рис. 1. Структурная схема процесса формирования в объекте сигналов поступающих на входы адаптивного подавителя помех

На рис. 2 приведена структурная схема предлагаемого АПП, который содержит два каскада.

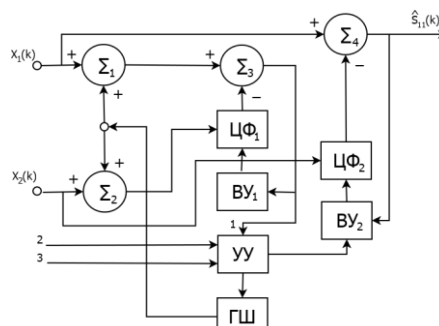


Рис. 2. Структурная схема двухкаскадного адаптивного подавителя помех в прерывистых и импульсных сигналах

На первом каскаде, образованном блоками суммирования $\Sigma_1, \Sigma_2, \Sigma_3$, цифровым фильтром ЦФ1, вычислительным устройством ВУ1, устройством управления УУ и генератором белого шума ГШ, реализуется первая ступень адаптивного подавления коррелированной помехи.

На втором каскаде, образованном блоком суммирования Σ_4 , цифровым фильтром ЦФ2 и вычислительным устройством ВУ2, реализуется вторая ступень – полное адаптивное подавление помехи с использованием вычисленных на первой ступени пауз.

В условиях высокого уровня мощности коррелированных помех, существенно превышающим уровень мощности полезного сигнала, определение пауз в полезном сигнале $s_{11}(k)$ по уровню мощности в сигнале $x_1(k)$ практически невозможно. В этих условиях для выделения пауз в полезном сигнале $s_{11}(k)$ предлагается производить декорреляцию [5] сигналов $x_1(k)$ и $x_2(k)$.

Декорреляция (отбеливание) осуществляется в первом каскаде АПП путем добавления в сигналы $x_1(k)$ и $x_2(k)$ на входах АПП белого шума с выхода ГШ. С помощью блоков Σ_1, Σ_2 белый шум суммируется с сигналами $x_1(k)$ и $x_2(k)$, что позволяет уменьшать долю мощности коррелированной составляющей помехи $s_2(k)$ в общей мощности аддитивной смеси сигналов $x_1(k)$ и $x_2(k)$. Освобождением от корреляции [5] достигается увеличение коэффициента подавления коррелированной помехи на первой ступени АПП.

Процесс декорреляции является итерационным и пошаговым. На каждом шаге ГШ по командам УУ увеличивает на определенную величину мощность шума. УУ анализирует изменения уровня мощности сигнала на выходе первой ступени АПП (выход блока Σ_3) и определяет наличие пауз.

Состояния пауз вычисляются УУ путем сравнения мощности сигнала, поступающего с выхода блока Σ_3 на первый вход УУ, с некоторым пороговой величиной, задаваемой на втором входе УУ. Эта пороговая величина определяет минимальный уровень мощности полезного сигнала $s_{11}(k)$, ниже которого сигнал считается отсутствующим, т.е. определяется состояние паузы.

На третьем входе УУ задается длительность шага итерации, определяемая величиной максимального временного интервала ожидания полезного сигнала $s_{11}(k)$ на первом входе АПП.

Если пауза определена, процесс итерации завершается и интервал паузы (моменты начала и завершения пауз) передаются в ВУ2 второго каскада АПП. Если пауза не определена, производится следующая итерация: ГШ увеличивается уровень мощности шума, УУ анализирует сигнал на выходе блока Σ_3 с целью определения пауз и т.д. Количество итераций ограничивается максимальным уровнем мощности коррелированной помехи, задаваемым для АПП.

На второй ступени производится полное подавление помех с использованием вычисленных на первой ступени пауз [2, 4], т.е. получение сигнала $\hat{s}_{11}(k)$, в котором помехи подавлены. Адаптация производится только в паузах полезного сигнала. Вне пауз сигнала адаптация запрещена и выделение сигнала $\hat{s}_{11}(k)$ путем вычитания из суммы $x_1(k)$ сигнала $x_2(k)$ производится с весовыми коэффициентами фильтра ЦФ2, вычисленными в конце предшествующей паузы и записанными в буферную память.

3. РЕЗУЛЬТАТЫ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Эффективность подавления помех предложенным двухкаскадным АПП подтверждается результатами компьютерного моделирования приведенными на рис.3. При моделировании применялся алгоритм адаптации Recursive Least Squares (RLS) [1], число весовых коэффициентов цифровых фильтров ЦФ1 и ЦФ2 равно 16.

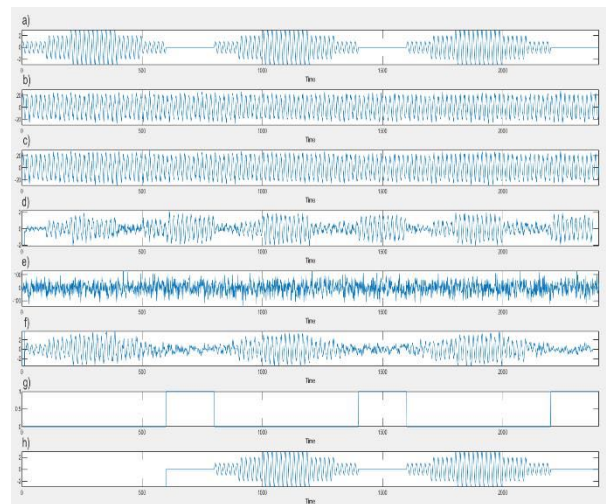


Рис. 3. Результаты компьютерного моделирования подавления помех двухкаскадным АПП

На рис. 3 приведены сигналы: (а) — прерывистый полезный сигнал $s_{11}(k)$; (б) — сумма сигналов помех $s_{21}(k)$ и $s_{31}(k)$, коррелированных с $s_{11}(k)$ и уровень мощности которых в 6 раз превышает уровень мощности полезного сигнала; (с) — сигнал $x_1(k)$ на первом входе АПП; (д) — пример практически неприемлемого подавления помех классическим АПП [1]; (е) — зашумленный ГШ при декорреляции сигнал на первом и втором входах АПП; (ф) — результаты подавления помех первой ступенью АПП (выход блока Σ_3) после 4 шагов процесса декорреляции; (г) — выделенные УУ интервалы пауз; (h) — сигнал $\hat{s}_{11}(k)$ на выходе второй ступени АПП. В результате подавления коррелированных помех сигнал $\hat{s}_{11}(k)$ равен сигналу $s_{11}(k)$ с приведенной погрешностью не превышающей 11%, что подтверждает эффективность применения предложенного двухкаскадного АПП.

4. ЗАКЛЮЧЕНИЕ

Разработаны алгоритм и реализующий его двухкаскадный адаптивный подавитель помех в прерывистых и импульсных сигналах. Применение декорреляции позволяет подавлять коррелированные помехи, уровень мощности которых значительно превышает уровень мощности полезных сигналов.

ЛИТЕРАТУРА

- [1] Джиган, В.И. Адаптивная фильтрация сигналов: Теория и алгоритмы / В.И. Джиган. — М.: Техносфера, 2013. — 528 с.
- [2] Засов, В.А. Адаптивный компенсатор помех в импульсных сигналах / В.А. Засов, М.В. Ромкин // Патент на изобретение RU №2736199 от 27.01.2020.
- [3] Засов, В.А. Адаптивный компенсатор помех в импульсных сигналах / В.А. Засов, М.В. Ромкин // Патент на изобретение RU №2735671 от 22.10.2019.
- [4] Zasov, V. Adaptive Cancellation of Interference in Intermittent and Pulse Signals / V. Zasov, M. Romkin // Data Science. Information Technology and Nanotechnology. Proc. of the Int. Conf. ITNT. IEEE Xplore, 2021. DOI: 10.1109/ITNT52450.2021.9649169.
- [5] Меркушев, А.В. Методы и алгоритмы разделения смеси сигналов. 1. Применение декорреляции и статистик второго порядка / А.В. Меркушев, Г.Ф. Малыгина // Научное приборостроение. — 2009. — Т. 19, № 2. — С. 90-103.

Решение задачи потокораспределения гидравлических сетей в условиях недостоверных исходных данных

А.О. Хуснутдинов

Новосибирский государственный технический университет

Новосибирск, Россия

evolext@gmail.com

В.С. Карманов

Новосибирский государственный технический университет

Новосибирск, Россия

karmanov@corp.nstu.ru

Аннотация—Работа посвящена решению проблемы расчета характеристик коммунальных сетей, в частности, выполнения гидравлических расчетов, в условиях отсутствия или искажения исходных данных. Предложена и исследована методика, позволяющая замещать пропуски в исходных данных на основе анализа данных по аналогичным или близким объектам.

Ключевые слова— коммунальные сети, водоснабжение, расчет потокораспределения, восстановление данных.

1. ВВЕДЕНИЕ

Согласно данным Министерства строительства и жилищно-коммунального хозяйства России, на момент весны 2020 года общая протяженность коммунальных сетей в России превышает 900 тыс. км., около 58% из них находится в изношенном состоянии [2]. Таким образом, существует необходимость в реконструкции и модернизации действующих коммунальных систем.

Для создания оптимальных конфигураций коммунальных систем, специалисты сферы ресурсообеспечения используют программные модели коммунальной инфраструктуры, что позволяет им автоматизировать большую часть работы, а также сократить число ошибок и неточностей в расчетах. При моделировании, примерно в 40% проектов, специалисты сталкиваются с важной проблемой – информация по характеристикам объектов коммунальных систем имеет пропуски (до 70% объектов) или искажения (до 100% объектов, в зависимости от срока службы системы), это приводит к тому, что расчеты по модели выполнить либо совсем невозможно, либо полученные результаты не будут соответствовать действительности.

Разработать методику расчета характеристик коммунальных систем при их моделировании, позволяющую решать задачу в условиях отсутствия или искажения исходных, а также провести исследование точности её программной реализации.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Термин «коммунальные сети» объединяет собой такие инженерные объекты, как сети водо-, тепло-, газоснабжения, водоотведения и другие. Несмотря на особенности, присущие каждому отдельному виду коммунальных сетей, для расчёта их характеристик используются похожие методы, учитывающие эти особенности. Дальнейшее изложение будет вестись на примере сетей водоснабжения, однако предлагаемый подход также применим и для сетей другого вида.

Расчет потокораспределения в сети является основным для коммунальных сетей, в частности, для

водопроводных сетей он заключается в определении величин расхода и потерь напора на всех участках сети.

Методы расчета базируются двух законах Кирхгофа [1]:

1) сумма расходов, втекающих в каждый узел равна нулю (или утечке);

2) сумма падений напора на всех участках, образующих замкнутый контур, равна нулю (или сумме действующих напоров).

Также применяется эмпирическая зависимость падения напора на участках сети от функции расхода вида:

$$\Delta H = Sf(q), \quad (1)$$

где ΔH – значения падений напора на всех участках, S – гидравлические сопротивления участков, q – значения расхода на участках сети, $f(q)$ – функция потерь от расхода, в частности, для водопроводных сетей имеет квадратичный вид.

Классическая задача распределения потока гидравлической сети может быть записана как [4]:

$$A^T H = Sf(q), \quad (2)$$

с начальными условиями вида

$$Aq = Q, \quad (3)$$

где A – матрица инцидентности графа сети без последней строки, H – напор на узлах сети, Q – расходы (утечки) в узлах.

При решении системы (2) могут быть неизвестны отдельные компоненты вектора расхода в узлах Q . В реальной системе величина расхода имеет вероятностный характер – её значения зависят от разных факторов, воздействие которых заранее предсказать невозможно.

Имея информацию по множеству аналогичных, с точки зрения технической конфигурации, коммунальных систем, можно применять статистические методы для замещения неизвестных значений характеристик объектов моделируемой сети. Например, использовать статистические критерии согласия эмпирического распределения неизвестной случайной величины с некоторым теоретическим для последующего выбора возможных замещающих значений.

Предлагается следующий алгоритм замещения неизвестной величины Q_i для некоторого i -узла:

1) подобрать данные о величине расхода на некоторых аналогичных, с точки зрения коммунальной конфигурации, рассматриваемому узлах;

2) с помощью некоторого критерия согласия определить по сформированной выборке теоретическое распределение $F(q)$ случайной величины расхода;

3) для неизвестной величины расхода, выбрать замещающее значение, определяемое видом выбранного распределения $F(q)$;

4) с помощью выбранного метода решения нелинейных систем уравнений выполнить гидравлический расчёт.

3. ПРИМЕР ВЫПОЛНЕНИЯ РАСЧЕТА

Исследуем точность предлагаемой методики решения задачи потокораспределения на типовой схеме сети водоснабжения, являющееся обобщением действующих систем водоснабжения небольших муниципальных образований. В качестве типовых объектов были выбраны данные по населенным пунктам Новосибирской области численностью до 1000 человек.

Средние характеристики сетей выбраны следующими: обобщенная схема водоснабжения состоит из одного источника – скважины; 80 потребителей водоснабжения с различными типовыми коммунальными конфигурациями (которые складываются из числа поливочных установок, наличия водонагревательного оборудования и числа кранов забора воды); а также 18 промежуточных узлов, обеспечивающих правильное функционирование водопроводной сети.

Объектам обобщенной сети зададим типовые значения основных гидравлических характеристик: расхода и напора – для узлов и коэффициентов сопротивления – для участков, далее выполним гидравлический расчёт значений расхода и потерь напора на участках сети.

Исследовать точность предложенной методики будем при разном количестве отсутствующих данных по потребителям – от 5 до 50%. В качестве замещающих значений могут использоваться различные характеристики модельных распределений, полученных после обработки выборки статистических данных – это могут быть теоретические и выборочные моменты, квантильные характеристики и другие величины. Для примера выберем математические ожидания случайной величины расхода, формулы для определения которых будут зависеть от выбранных законов распределения.

Полученные результаты сравним с найденными ранее точными результатами величин расхода на участках по значениям максимальной абсолютной погрешности и максимальной относительной погрешности среди всех участков. Графики изменения максимальной абсолютной и максимальной относительной погрешностей для различных пар выбранных распределений приведены на Рис. 1 и Рис. 2 соответственно.

Из полученных графиков следует вывод, что даже при более точном выборе вида распределения неизвестной случайной величины, неточность результатов может достигать до 15% и более. Следовательно, задача сбора качественных данных и их обработка современными математическими методами продолжает оставаться актуальной.

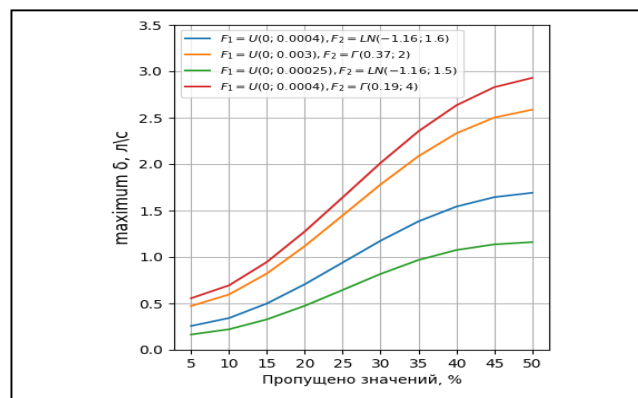


Рис. 1. Изменение максимальной абсолютной погрешности рассчитанных значений расхода

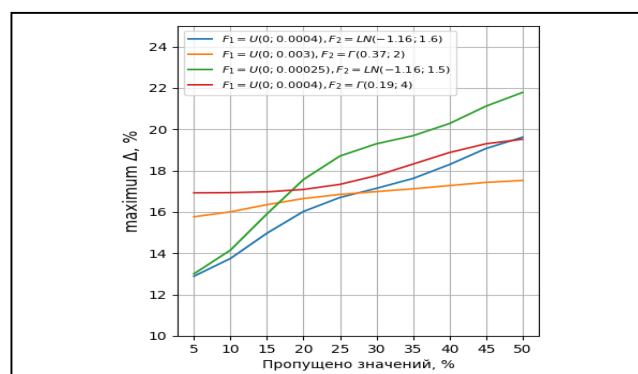


Рис. 2. Изменение максимальной относительной погрешности рассчитанных значений расхода

4. ЗАКЛЮЧЕНИЕ

В представленной работе была предложена методика расчета неизвестных характеристик объектов коммунальных систем при их моделировании в условиях отсутствия части исходных данных, которая была реализована в виде программного модуля и исследована на типовой задаче потокораспределения в водопроводной сети.

На следующем этапе работы стоит рассмотреть иные методы замещения неизвестных характеристик объектов и провести аналогичное исследование погрешности результатов гидравлического расчёта.

ЛИТЕРАТУРА

- [1] Антропов, Н.Р. Идентификация параметров гидравлического сопротивления модели гидравлической цепи / Н.Р. Антропов, Е.Д. Агафонов // Сибирский журнал науки и технологий. – 2017. – Т. 18, № 3. – С. 492-498.
- [2] Минстрой сообщил об износе коммунальных сетей в России на 58% [Электронный ресурс]. – Режим доступа: <https://www.interfax.ru/business/700998> (01.02.2022).
- [3] Якшин, С.В. Аналитический метод решения задачи потокораспределения тепловой сети. Вестник Иркутского государственного технического университета / С.В. Якшин. – 2021. – Т. 25, № 1. – С. 80-96. DOI: 10.21285/1814-3520-2021-1-80-96.
- [4] Korelstein, L. Modernization of Todini Global Gradient Algorithm for hydraulic analysis of networks with choked flow / L. Korelstein // E3S Web of Conferences. – 2020. – № 219. – P. 01003.
- [5] Khan, W.A. Numerical and simulation analysis comparison of hydraulic network problem base on higher-order efficiency approach / W.A. Khan // Alexandria Engineering Journal. – 2021. – Vol. 60. – P. 4889-4903.

Децентрализованный алгоритм обнаружения изменения режима движения объекта по данным мультисенсоров

А.В. Цыганов
Ульяновский государственный
педагогический университет
им. И.Н. Ульянова
Ульяновск, Россия
andrew.tsyanov@gmail.com

Ю.В. Цыганова
Ульяновский государственный
университет
Ульяновск, Россия
tsyganovajv@gmail.com

А.В. Голубков
Ульяновский государственный
педагогический университет
им. И.Н. Ульянова
Ульяновск, Россия
kr8589@gmail.com

Аннотация—В работе рассмотрена задача обнаружения изменения режима движения объекта в неизвестный момент времени по данным мультисенсоров. Предложено новое решающее правило, для вычисления которого используется децентрализованный алгоритм дискретной фильтрации. Алгоритм реализован в системе MATLAB. Результаты компьютерного моделирования подтверждают работоспособность предложенного решения.

Ключевые слова— децентрализованный алгоритм обработки данных, распределенная фильтрация Калмана, последовательное решающее правило, модели движения объекта.

1. ВВЕДЕНИЕ

Мультисенсорные сети широко применяются в различных областях науки и техники, например, в системах управления с целью улучшения качества получаемых оценок состояния изучаемого объекта или системы за счет синтеза данных с различных сенсоров. В связи с этим актуальной является разработка новых и модернизация имеющихся алгоритмов оптимального оценивания в мультисенсорных сетях.

Одним из наиболее популярных методов оценивания на протяжении многих десятилетий является фильтр Калмана, который нашел свое применение и в мультисенсорных системах. Обзор методов распределенной калмановской фильтрации можно найти, например, в [1, 2].

В [3] рассматривается задача мультисенсорного оценивания параметров модели объекта с несколькими возможными режимами движения. Предложен децентрализованный мультисенсорный алгоритм оценивания вектора состояния модели движения на основе децентрализованного алгоритма калмановской фильтрации.

В [4] предложен гарантированный по вероятностям ошибок первого и второго рода метод обнаружения факта изменения режима движения объекта в заранее неизвестный момент времени. Решение основано на представлении траектории движения объекта гибридной стохастической моделью, применении последовательного решающего правила и алгоритма калмановской фильтрации. Отличие данного метода в том, что решение принимается на ограниченном множестве значений функции отношения правдоподобия.

Целью данной работы является развитие результатов, полученных в [3] и [4], а именно: построение нового

децентрализованного алгоритма обнаружения изменения режима движения объекта в неизвестный момент времени по измерительным данным, получаемым от сети сенсоров.

2. ДЕЦЕНТРАЛИЗОВАННЫЙ АЛГОРИТМ ОБНАРУЖЕНИЯ ИЗМЕНЕНИЯ РЕЖИМА ДВИЖЕНИЯ ОБЪЕКТА

Рассмотрим сеть сенсоров с полностью связанной топологией, состоящую из N узлов, в которой каждый узел имеет возможность вычислять собственные оценки вектора состояния модели движения объекта. Измерения и оценки, получаемые в узлах, будем называть локальными.

Предположим, что каждый узел имеет одинаковый набор моделей объекта, описывающих возможные режимы его движения

$$x(k) = \Phi_p x(k-1) + B_p u(k-1) + G_p w_p(k-1),$$

а локальные измерения описываются моделью

$$z_i(k) = H_{ip} x(k) + v_{ip}(k), \quad k=1,2,\dots,$$

где $p=0,1,\dots,M-1$, $i=1,\dots,N$, M — количество возможных режимов движения, N — количество узлов, $x(k)$ — вектор состояния объекта, $u(k)$ — вектор управления, $z_i(k)$ — вектор измерений, $w_p(k)$ — гауссовский шум в объекте, $v_{ip}(k)$ — гауссовский шум в измерителе.

Ключевая идея децентрализованного фильтра Калмана состоит в возможности выразить глобальные оценки вектора состояния через локальные [2]. Локальные обновления величин фильтра вычисляются в каждом узле и передаются всем остальным узлам.

Предположим, что в некоторый заранее не известный момент времени может произойти изменение режима движения объекта. Задача заключается в обнаружении данного факта. В [4] было предложено последовательное решающее правило, позволяющее принять решение об изменении режима движения на ограниченном множестве значений функций правдоподобия. В данной работе предлагается решающее правило с новым выражением для вычисления функции правдоподобия (для двух режимов движения):

$$\lambda_k = \frac{1}{k} \sum_{j=1}^k \psi_j^1(k),$$

где значения $\psi_j^1(k)$ вычисляются на основе оценок, получаемых фильтром Калмана.

Отличие от [4] заключается в том, что для вычисления λ_k мы используем децентрализованный алгоритм фильтрации [3] на основе информационной формы фильтра Калмана, который позволяет получить все требуемые для расчетов величины в каждом узле по данным мультисенсоров. Таким образом, новое решающее правило с вычислением λ_k позволяет динамически контролировать процесс изменения режима движения объекта в каждом узле мультисенсорной сети.

3. ПРИМЕР

Рассмотрим траекторию движения объекта, приведенную на рис. 1. Объект начинает движение из точки с координатами (0,0) с вектором скорости (-2,2). Сначала он в течение 40 дискретных моментов совершает равномерное круговое движение по часовой стрелке с радиусом 3, а затем равномерное круговое движение против часовой стрелки с радиусом 5.

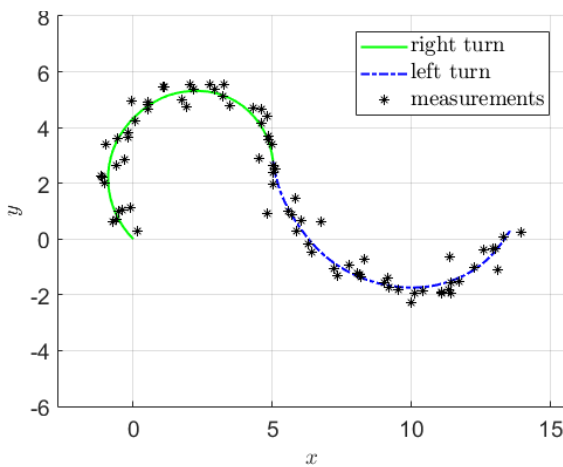


Рис. 1. Траектория объекта

На рис. 2 приведен график функции правдоподобия λ_k , вычисляемой в каждом узле.

4. ЗАКЛЮЧЕНИЕ

В работе построен децентрализованный алгоритм обнаружения изменения режима движения объекта по данным мультисенсоров. Предложено новое выражение для вычисления функции правдоподобия в децентрализованном информационном фильтре.

Результаты компьютерного моделирования в среде MATLAB подтвердили работоспособность предложенного решения.

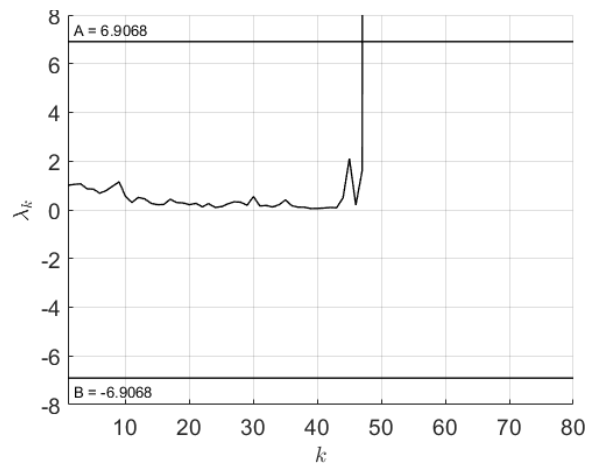


Рис. 2. График λ_k

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ и Правительства Ульяновской области в рамках научных проектов № 18-41-732001 р_мк, № 18-41-732003 р_мк.

ЛИТЕРАТУРА

- [1] Mahmoud, M.S. Distributed Kalman filtering: a bibliographic review / M.S. Mahmoud, H.M. Khalid // IET Control Theory Appl. – 2013. – Vol. 7(4). – P. 483-501. DOI: 10.1049/iet-cta.2012.0732.
- [2] Rao, B. Fully decentralized algorithm for multisensor Kalman Filtering / B. Rao, H. Durrant-Whyte // IEEE Proceedings. – 1991. – Vol. 138(5). – P. 413-420.
- [3] Golubkov, A.V. Decentralized multisensor estimation of motion parameters of an object moving along a complex trajectory / A.V. Golubkov, A.V. Tsyganov, Yu.V. Tsyganova, I.O. Petrishchev // Journal of Physics: Conference Series. – 2019. – Vol. 1368. – P. 042041. DOI: 10.1088/1742-6596/1368/4/042041.
- [4] Golubkov, A.V. Algorithm for detecting a change in the motion mode of an object moving along a complex trajectory / A.V. Golubkov, A.V. Tsyganov, I.O. Petrishchev // Journal of Physics: Conference Series. – 2021. – Vol. 1745. – P. 012115. DOI: 10.1088/1742-6596/1745/1/012115.

Использование нейросетевой модели для исследования нефтяных объектов

А.А. Минигалиев
Казанский национальный
исследовательский технический
университет им. А.Н. Туполева –
КАИ
Казань, Россия
ainurminigaliev4@gmail.com

И.М. Кычкин
Казанский национальный
исследовательский технический
университет им. А.Н. Туполева –
КАИ
Казань, Россия
vip.kychkin@gmail.com

В.В. Мокшин
Казанский национальный
исследовательский технический
университет им. А.Н. Туполева –
КАИ
Казань, Россия
vladimir.mokshin@mail.ru

Аннотация—Рассмотрен вопрос моделирования функционирования месторождения в нефтяной промышленности. В работе приведено описание процесса обучения нейросетевой модели. Приведены графические материалы по результатам обучения и тестирования нейросети. Для нахождения параметров рассматриваемой производственной системы, обеспечивающих максимальную эффективность работы, авторы реализовали сбор статистики.

Ключевые слова—*MatLAB, глубокое обучение, нейросеть, производство, добыча нефти.*

1. ВВЕДЕНИЕ

В наши дни в нефтяную промышленность все чаще интегрируются информационные технологии. Это позволяет значительно повысить эффективность добычи ископаемых и сокращать необходимые для этого ресурсы. Более того, появляются широкие возможности по сбору статистики, что в свою очередь позволяет проводить моделирование, осуществляя концепцию «бережливого производства». Данный подход подразумевает постоянное стремление к устранению всех видов потерь и предполагает вовлечение в процесс оптимизации производства каждого сотрудника и максимальную ориентацию на потребителя, что считается сегодня одной из самых эффективных моделей развития компаний [1].

Целью исследования является анализ нейросетевой модели исследования нефтяных объектов для мониторинга работы скважин на основе интеллектуальных технологий и использования глубокого обучения. Задачами является: моделирование процесса режима работы скважин; анализ перечня переменных необходимых для исследования модели; анализ обучения и тестирование нейросети.

Публикация соответствует тематике секции «Науки о данных», т. к. в работе были получены необходимые оптимизационные данные для формирований дальнейших рекомендаций по управлению работой скважинами и использования для вычисления основных статистических характеристик, регрессионного, корреляционного анализа, рассмотрения взаимосвязи производственно-технических показателей между собой.

2. ВЫБОР СОВОКУПНОСТИ РЕЗУЛЬТАТИВНЫХ ПОКАЗАТЕЛЕЙ И ВЛИЯЮЩИХ НА НИХ ФАКТОРОВ

Для исследования влияния производственно-технических факторов скважины по результатам

экспертных оценок выбираются результативные показатели состояния скважины – $y_i, i = 1, k$. В качестве влияющих на них факторов выбираем производственно-технические факторы – $x_i, i = 1, m$. Обозначаем все переменные (признаки), отобранные для исследования совокупностью – $v_j; j = 1, m + k$. Перечень переменных приведен в таблице 1.

ТАБЛИЦА 1. ПЕРЕЧЕНЬ ПЕРЕМЕННЫХ, ОТОБРАННЫХ ДЛЯ ИССЛЕДОВАНИЯ

№	Имя	Описание
1	x_1	Дебит по жидкости в м ³ /сут. (Qж_м3_сут)
2	x_2	Дебит по жидкости в т/сут. (Qж_т_сут)
3	x_3	Дебит по нефти в т/сут. (Qн_т_сут)
4	x_4	Обводненность (Обв)
5	x_5	Обводненность вес (Обв_вес)
6	x_6	Плотность воды (Пл_воды)
7	x_7	Жидкость в м ³ /мес. (Жид_м3_мес)
8	x_8	Жидкость в т/мес. (Жид_т_мес)
9	x_9	Нефть в т/мес. (Неф_т_мес)
10	x_{10}	Вода в т/мес. (Вода_т_мес)
11	x_{11}	Время работы (Т_раб)
12	x_{12}	Время простоя (Т_пр)
13	x_{13}	Время нак. (Т_нак)
14	x_{14}	Жидкость в м ³ /год (Жид_м3_год)
15	x_{15}	Жидкость в т/год (Жид_т_год)
16	x_{16}	Нефть в т/год (Неф_т_год)
17	x_{17}	Вода т/год (Вода_т_год)
18	x_{18}	Время работы в год (Траб_год)
19	x_{19}	Эксплуатационная жидкость в м ³ (Жид_м3_экс)
20	x_{20}	Эксплуатационная жидкость в т (Жид_т_экс)
21	x_{21}	Эксплуатационная нефть в т (Неф_т_экс)
22	x_{22}	Эксплуатационная вода в т (Жид_т_экс)
23	x_{23}	Код насоса (Код_нас)
24	x_{24}	Пластовое давление (P_пл)
25	x_{25}	Забойное давление (P_заб)
26	y_1	Категория рентабельности (Кат_рент)
27	y_2	Маржинальный доход в руб./сут. (Марж_дох_руб_сут)
28	y_3	Себестоимость 1 тонны нефти (Себест_1т_н)
29	y_4	Цена 1 тонны нефти (Цена_1т_н)
30	y_5	Затраты в руб./мес. (Затраты_руб_мес)
31	y_6	Мероприятия (Мероприят)

3. ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ МЕТОДОМ ЛЕВЕНБЕРГА-МАРКВАРДА

Алгоритм Левенберга-Марквардта предназначен для оптимизации параметров нелинейных регрессионных моделей. Предполагается, что в качестве критерия оптимизации используется среднеквадратичная ошибка модели на обучающей выборке. Алгоритм заключается в последовательном приближении заданных начальных значений параметров к искомому локальному оптимуму. Когда данные сильно различаются единицами измерения и/или диапазоном значений, перед применением того или иного метода анализа. Стандартизация данных осуществляется по формуле:

$$x_{si} = \frac{x_i - \bar{x}_i}{\sigma_i^2}$$

Далее проводим загрузку данных в MatLAB и обучаем на наших входных данных нейронную сеть методом Левенберга-Марквардта. Для нашей задачи мы спроектировали динамическую нейронную сеть, которая представляет собой двухслойную сеть прямой связи с сигмоидной передаточной функцией в скрытом слое и линейной передаточной функцией в выходном слое. Эта сеть также использует отводимые линии задержки для хранения предыдущих значений и последовательности.

Используя обученную нейронную сеть, построили график изменения ошибки обучения по Y1, Y2, Y3, а также спрогнозировали результативные показатели. На рисунке изображены результаты обучения нейронной сети для Y1 и значение коэффициента детерминации. На рисунке 2 представлен график изменения ошибки обучения Y1.



	Observations	MSE	R
Training	208	5.8227	0.9252
Validation	45	45.6943	0.5973
Test	45	51.2504	0.5516

Рис. 1. Результаты обучения нейронной сети для Y1

Для того, чтобы спрогнозировать результативные показатели, необходимо сначала обучить нейронную сеть, используя обучающую выборку, которая предназначена для построения классификаторов, а потом использовать тестовую выборку, предназначенную для оценки качества работы сети, на уже обученной ранее нейронной сети. На рисунке 3 проиллюстрирована разница между целевыми данными и полученной аппроксимирующей кривой.

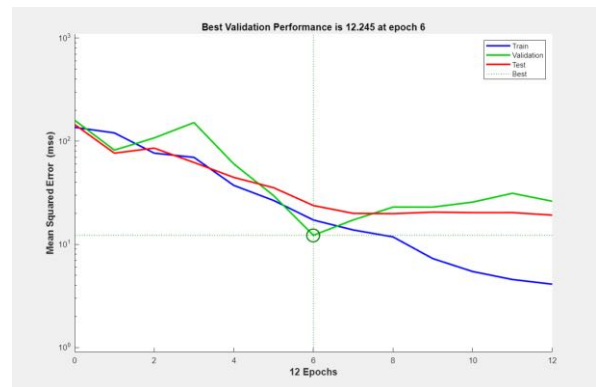


Рис. 2. Изменение ошибки обучения (Y1)

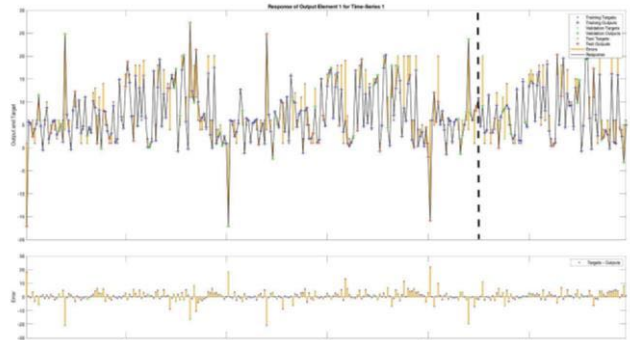


Рис. 3. График функций обученной нейронной сети для Y1

3. ЗАКЛЮЧЕНИЕ

Таким образом, результатом данной работы является обученная нейронная сеть и ее анализ для исследования нефтяных объектов. В ходе экспериментов по моделированию на основе статистики была осуществлена настройка модели, были получены оптимизационные данные, необходимые для обучения нейронной сети. Были построены графики по оценке ошибки обучения и спрогнозированы результативные показатели. Тем самым, можно сказать, что полученные данные можно использовать для дальнейшей работы, анализа и внедрения.

ЛИТЕРАТУРА

- [1] Чернова, В.А. Концепция бережливого производства: неуклонное сокращение потерь / В.А. Чернова, И.Т. Агеев // Молодой ученый. – 2016. – № 26(130). – С. 407-410.
- [2] Мокшин, В.В. Анализ динамической модели «Оптимизация дорожного движения на участке дороги» в среде STATISTICA и Anylogic / В.В. Мокшин, А.П. Кирпичников, Р.Р. Шириязданов, Ш.А. Шайхутдинов // Вестник технологического университета. – 2019. – Т. 22, № 5.
- [3] Дискретно-событийное моделирование. Anylogic. [Электронный ресурс]. – Режим доступа: <https://www.anylogic.ru/use-of-simulation/discrete-event-simulation/> (23.01.22)
- [4] Мокшин, В.В. Имитационное моделирование процесса сортировки и выдачи багажа в аэропорту / В.В. Мокшин, А.П. Кирпичников, О.К. Марусина, Д.М. Зиятдинов, В.А. Евлампьев, Р.Р. Шириязданов. // Вестник технологического университета. – 2019. – Т. 22, № 6.
- [5] Федотова, В.С. Технологии имитационного моделирования в системе AnyLogic / В.С. Федотова // Царскосельские чтения. – 2013. – № XVII [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/n/tehnologii-imitatsionnogo-modelirovaniya-v-sisteme-anylogic> (04.02.2022).

Технология применения метода поиска информативных признаков для снижения размерности признакового пространства в задаче классификации областей натуральных гиперспектральных изображений

М.И. Хотилин

Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
khotilin.mi@ssau.ru

Аннотация—В данной статье описывается процесс применения метода отбора информативных признаков области гиперспектрального изображения для осуществления классификации. Описаны методы и алгоритмы поиска признаков принадлежности к определенной области, технология их применения. Указаны дальнейшие перспективы развития алгоритма.

Ключевые слова— гиперспектральные изображения, дискриминантный анализ, классификация, отбор признаков, кластеризация, Python, снижение размерности, информативные признаки.

1. ВВЕДЕНИЕ

Гиперспектральные изображения – это трёхмерный массив данных, который включает в себя пространственную информацию об объекте, дополненную спектральной информацией по каждой пространственной координате [1]. Обработка и анализ гиперспектральных изображений и их областей получили широкую популярность среди тематик в области обработки изображений и компьютерного зрения. В рамках данной работы рассматривается технология применения метода поиска информативных признаков гиперспектрального изображения для задачи кластеризации, на примере отдельно взятой его области.

2. ПОСТАНОВКА ЗАДАЧИ

Выполнение обработки RGB изображений классическими методами, их классификация занимает сравнительно небольшое время и может выполняться практически на любом устройстве, в том числе и носимом. В то же время, ввиду ряда особенностей, обработка гиперспектральных изображений, будет требовать значительных вычислительных ресурсов. Например, для классификации одной области изображения гиперспектрального размером 10×10 пикселей, количество необходимых совокупных яркостных и текстурных признаков составляет 86878.

Актуальность данной работы заключается в создании технологии, основанной на методе поиска информативных признаков изображения, позволяющей значительно снизить временные и аппаратные ресурсы, используемые в процессе анализа гиперспектральных изображений.

Весь процесс данной работы можно разделить на последовательно выполняемые этапы. На первом этапе посредством использования модуля предобработки выделяется содержащая различные объекты область изображения, используемая для исследования и обработки. Далее, из гиперкуба исследуемой области выделяется набор всех двумерных сечений плоскостями. Данный этап необходим ввиду того, что заранее не представляется возможным сказать, какие именно слои являются значимыми, а также потому, что применение существующих методов и средств вычисления признаков является ресурсозатратным.

Для исследования и обработки полученных выше сечений было принято решение использовать хорошо зарекомендовавшее себя ПО «MaZda» [4,6], позволяющее произвести расчет различных групп признаков, а также высокоуровневый язык программирования Python. В результате работы данного программного продукта получаем набор текстурных и яркостных признаков, используемый в дальнейшем.

Исследуемые изображения могут содержать значительное количество шумовых компонентов, ввиду чего необходимо произвести обработку, позволяющую сгладить их. Далее, явно задавая количество кластеров, можно произвести разделение исследуемых данных, получая, таким образом, наборы данных (признаков) для обучения, сгруппированных по определенным критериям. Полученные наборы признаков являются значительными по объему и содержат данные, которые могут не нести значимой информации, важной при классификации. В связи с этим, необходимо произвести сокращение размерности и поиск признаков, являющихся информативными.

Для поиска информативных признаков могут применяться различные алгоритмы. В рамках данной работы использовался метод последовательного добавления признаков. Далее, используя различные алгоритмы классификации (LDA, SVM), можно произвести классификацию полученных данных.

3. ПРАКТИЧЕСКОЕ ИССЛЕДОВАНИЕ И ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

В качестве исходного рассматривалось изображение, размером 1311×3411 пикселей с количеством

спектральных каналов 242. Для исследования описанного выше алгоритма, посредством модуля предобработки, реализованного на Python, из исходного изображения вырезались небольшие области.

Далее, с помощью того же модуля, полученный гиперкуб изображения - массив яркостей - разделялся на двумерные по каждой паре координат и полученные сечения гиперкуба использовались для дальнейшего получения набора признаков, как яркостных, так и текстурных, извлеченных с помощью ПО «MaZda». Для небольших областей изображения, размером 10×10 пикселей, количество совокупных яркостных и текстурных признаков составляет 86878 признаков (24200 яркостных признаков + 62678 текстурных).

Следующим этапом, используя методы удаления шумов (медианный фильтр) и повышения резкости изображения (unsharp masking), было произведено «сглаживание» изображения и повышение резкости границ объектов рассматриваемой области изображения. В результате были получены текстурные и яркостные признаки. Далее путем варьирования величины количества кластеров (2, 3, 4, 5), была проведена группировка полученных данных в кластеры. После проведения данного этапа необходимо сократить размерность признакового пространства ввиду того, что вычисления по всем имеющимся признакам могут занимать значительное время и значительные вычислительные ресурсы.

Для сокращения размерности использовался метод, описанный ранее автором в [6], состоящий из совместного использования линейного дискриминантного и корреляционного анализа, а для поиска информативных признаков использовался метод последовательного добавления признаков. В результате удалось сократить размерность с 86878 признаков до 39, то есть более чем в 2000 раз. При этом размер используемой в процессе вычислений оперативной памяти был уменьшен с 24 Gb до 6Gb.

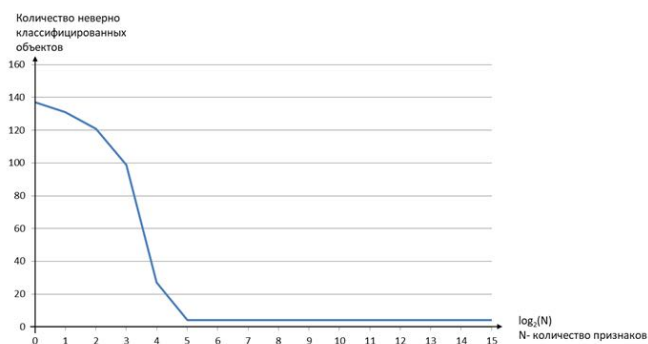


Рис. 1. Зависимость количества неверно классифицированных гиперспектральных элементов от размера предобработанного массива признаков

Далее к полученным данным применяются различные алгоритмы классификации: LDA, SVM и др. В качестве исходных данных для проверки работы описанного в статье алгоритма, можно использовать предварительно

размеченные изображения, удаляя их разметку и подавая неразмеченные данные на вход.

Оценивая параметры работы данных алгоритмов, можно выбрать наиболее эффективный по точности классификации или времени его работы. В итоге получаем паттерн обработки гиперспектральных изображений, обладающий меньшими ресурсными требованиями, по сравнению с классическими методами, позволяющий использовать его для различных целей, например, на мобильных устройствах или беспилотных летательных аппаратах.

4. ЗАКЛЮЧЕНИЕ

Нахождение признаков, определяющих однозначно принадлежности областей изображения к определенному классу, является одной из значимых задач классификации и обработки изображений. Существующие методы обработки изображений, определения их признаков и алгоритмы классификации отлично работают с относительно небольшими объемами исходных данных. Вычисления, описанные в данной работе, являлись небольшими и были выполнены на персональном компьютере. Обработка больших массивов исходных изображений занимает значительное время и вычислительные ресурсы.

В настоящее время ведется работа над исследованием возможности построения алгоритма, способного значительно эффективнее справляться с задачей поиска информативных признаков областей гиперспектральных изображений и их классификации, при сохранении точности данных процессов, используя программные средства для обработки данных больших объемов.

ЛИТЕРАТУРА

- [1] Zimichev, E.A. Spectral-spatial classification with k-means++ partitional clustering / E.A. Zimichev, N.L. Kazanskiy, P.G. Serafimovich // *Computer Optics*. – 2014. – Vol. 38(2). – P. 281-286. DOI: 10.18287/0134-2452-2014-38-2-281-286.
- [2] Kazanskiy, N.L. Simulation of hyperspectrometer on spectral linear variable filters / N.L. Kazanskiy, S.I. Kharitonov, S.N. Khonina, S.G. Volotovskiy, Yu.S. Strelkov // *Computer Optics*. – 2014. – Vol. 38(2). – P. 256-270. DOI: :10.18287/0134-2452-2014-38-2-256-270.
- [3] Strzelecki, M. A software tool for automatic classification and segmentation of 2D/3D medical images / P. Szczypinski, A. Materka, A. Klepaczko // *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. – 2013. – Vol. 702. – P. 137-140. DOI: 10.1016/j.nima.2012.09.006.
- [4] Khotilin, M. Classification of objects of natural hyperspectral images / M. Khotilin, N. Kravtsova, I. Rytsarev, A. Kupriyanov // *International Conference on Information Technology and Nanotechnology (ITNT)*. – 2020. – P. 1-3. DOI: 10.1109/ITNT49337.2020.9253254.
- [5] Goncharova, E.F. Greedy algorithms of feature selection for multiclass image classification / E.F. Goncharova, A.V. Gaidel // *CEUR Workshop Proceedings*. – 2018. – Vol. 2210. – P. 38-46. DOI: 10.18287/1613-0073-2018-2210-38-46.
- [6] Khotilin, M. The technology of constructing an informative feature of a natural hyperspectral image area for the classification problem / M. Khotilin // *International Conference on Information Technology and Nanotechnology (ITNT)*. – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649178.

Анализ алгоритмов поиска подстроки в тексте с учетом их практического применения на основе экспериментов

В.А. Михайлов
Казанский (Приволжский) федеральный университет
Казань, Россия
v.mikhailov23@gmail.com

Э.Н. Мифтахов
Башкирский государственный университет
Уфа, Россия
promif@mail.ru

Т.А. Михайлова
Башкирский государственный университет
Уфа, Россия
t.a.mihailova@yandex.ru

С.А. Мустафина
Башкирский государственный университет
Уфа, Россия
mustafina_sa@mail.ru

Аннотация—В статье приведены результаты исследования детерминированных алгоритмов решения задачи поиска подстроки в строке – простейший последовательный алгоритм поиска, алгоритм Рабина-Карпа, алгоритм Кнута-Морриса-Пратта и алгоритм Бойера-Мура. Посредством проведения анализа асимптотических оценок сложности алгоритмов построен вероятностный алгоритм поиска и осуществлена его программная реализация. Для моделирования практического применения рассматриваемых алгоритмов проведены вычислительные эксперименты, результаты которых отражают способы применения каждого алгоритма.

Ключевые слова— поиск подстроки в тексте, детерминированный алгоритм, вероятностный алгоритм поиска, оценка сложности, простейший последовательный алгоритм поиска, алгоритм Рабина-Карпа, алгоритм Кнута-Морриса-Пратта, алгоритм Бойера-Мура.

1. ВВЕДЕНИЕ

Одной из популярных задач поиска информации является поиск подстроки в тексте. Данный класс задач аналогичен поиску «иголки в стоге сена» и описывается следующим образом. Пусть задан текст t и шаблон (строка) p . Необходимо определить, входит ли некоторый заданный шаблон символов в исходный текст. В случае вхождения – найти массив номеров символов, начиная с которого p содержится в t , в противном случае – вывести пустой массив.

Описанная задача встречается на практике очень часто как функция во множестве часто используемых приложений, таких как текстовые редакторы, браузеры, поисковые системы, СУБД, системы определения плагиата, системы анализа хакерских атак и т.п.

Для решения данной постановки существует множество алгоритмов. Следовательно, актуальной является задача выбора алгоритма поиска подстроки в строке, для решения которой целесообразно провести их экспериментально-сравнительный анализ посредством программной реализации детерминированных и вероятностного алгоритмов поиска [1].

2. ОЦЕНКА СЛОЖНОСТИ

В рамках исследования были изучены детерминированные алгоритмы поиска подстроки:

- простейший последовательный алгоритм поиска;
- алгоритм Рабина-Карпа [2, 3];
- алгоритм Кнута-Морриса-Пратта [2, 3];
- алгоритм Бойера-Мура [4].

Первоначально в результате, приведенном в таблице 1, оценки сложности на модели однопроцессорной машины с произвольным доступом к памяти (как наиболее близкой к работе современных процессоров) был выбран и построен алгоритм на основе выполнения алгоритма Рабина-Карпа. Так как другие алгоритмы имели в качестве недостатка проблему плохой модифицируемости получения универсальной для разных шаблонов префикс-функции и таблицы хороших суффиксов, так как они полностью зависят от содержания текста шаблона.

ТАБЛИЦА I. Оценки временной сложности детерминированных алгоритмов поиска подстроки

Algorithm	Временная сложность		
	предварительных вычислений	фазы сравнения	сравнения в худшем случае
Простейший алгоритм поиска	0	$O(m(n-m+1))$	$O(n^2)$
Алгоритм Кнута-Морриса-Пратта	$O(m)$	$O(n)$	$O(n+m)$
Алгоритм Бойера-Мура	$O(m + \Sigma)$	$O(n)$	$O(n+m+ \Sigma)$
Алгоритм Рабина-Карпа	$O(m)$	$O(n)$	$O(m(n-m+1))$

Также стоит отметить, что, хотя и алгоритму Рабина-Карпа свойственно возникновение коллизии хешей разных строк (что может иногда приводить к неверному результату), он решает проблему неверного ответа путем посимвольной перепроверки шаблона и рассматриваемого фрагмента текста, чьи хеши совпадают [4, 5]. Однако временная сложность в худшем случае может стать не лучше средней сложности простейшего алгоритма последовательного поиска. Поэтому в качестве основной идеи построения вероятностного алгоритма замена посимвольной проверки на повторный вызов алгоритма, используя другую хеш-функцию. Для упрощения был использован кольцевой полиномиальный

хеш, где предварительно случайным образом выбиралось полиномиальное основание. Это позволяет уменьшить временную сложность алгоритма в худшем случае до $O(n+m)$.

3. РЕАЛИЗАЦИЯ АЛГОРИТМОВ И ПРОВЕДЕНИЕ ЭКСПЕРИМЕНТА

В процессе исследования алгоритмы были программно реализованы на языке программирования Java и протестированы на ЭВМ со следующей конфигурацией:

- процессор Intel Core i5-9300h, 4x2.4 GHz;
- оперативная память DDR4 16GB 2666 MHz (двухканальный режим памяти).

При этом эксперименты были проведены повторно несколько раз, чтобы учесть зависимость результатов от характеристик и загрузки процессора фоновыми приложениями, чтения-записи информации в ОЗУ. В качестве конечного результата использовалось среднее полученных результатов выполнения соответствующего алгоритма.

Следует также отметить, что во всех экспериментах выполнялся поиск по всем позициям строк, поэтому эксперименты близки к наихудшим случаям, поскольку все алгоритмы должны проверять все возможные позиции подстрок на соответствие шаблонам.

Вычислительные эксперименты были проведены в форме поиска одного и нескольких шаблонов в неупорядоченной базе данных, содержащей более миллиарда строк, фрагмент которого приведён в таблице 2, и поиска одного и нескольких шаблонов в тексте книги – первый том произведения Толстого Л.Н. «Война и мир», содержащий 1274973 символа.

ТАБЛИЦА II. ФРАГМЕНТ НЕУПОРЯДОЧЕННОЙ БАЗЫ ДАННЫХ

8536	22802	3212	55444	2016-12-02	17:52:32+03
672360	64356	139276	7254	2012-04-25	22:37:38+06
96995	32916	234170	1203	2004-12-31	05:54:51+04
658474	19236	22357	3379	2012-02-11	14:15:29+02

Результаты экспериментально-сравнительного анализа приведены в таблице 3 и 4. Они показали, что алгоритмы Бойера-Мура и Кнута-Морриса-Пратта имеют значительные преимущества в одиночном поиске и часто используются в повседневных приложениях (текстовые редакторы, поиск в странице браузера и т.д.), однако плохо модифицируются для параллельных и вероятностных вычислений, что даёт преимущество алгоритму Рабина-Карпа там, где необходима частая и параллельная обработка фрагментов в часто использующихся данных (системы антиплагиата, системы учёта отзывов, анализ хакерских атак). Простейший алгоритм последовательного поиска полезен для ознакомления и практически возможен в использовании поиска в документе небольшого объёма.

ТАБЛИЦА III. СРЕДНЕЕ ВРЕМЯ ВЫПОЛНЕНИЯ ЭКСПЕРИМЕНТОВ ПОИСКА В НЕУПОРЯДОЧЕННОЙ БАЗЕ ДАННЫХ

Алгоритм	Среднее время выполнения (мс)			
	Без учета предварительных вычислений		Без учета предварительных вычислений	
	1 шаблон	1 шаблон	1 шаблон	1 шаблон
Простейший алгоритм поиска	196	654	196	654
Алгоритм Кнута-Морриса-Пратта	132	350	154	467
Алгоритм Бойера-Мура	89	359	113	484
Алгоритм Рабина-Карпа	57	240	757	944
Вероятностный алгоритм	53	179	3653	3838

ТАБЛИЦА IV. СРЕДНЕЕ ВРЕМЯ ВЫПОЛНЕНИЯ ЭКСПЕРИМЕНТОВ ПОИСКА В КНИГЕ

Алгоритм	Среднее время выполнения (мс)			
	Без учета предварительных вычислений		С учетом предварительных вычислений	
	1 шаблон	5 шаблонов	1 шаблон	5 шаблонов
Простейший алгоритм поиска	14	325	14	325
Алгоритм Кнута-Морриса-Пратта	~2.52	11	4	18
Алгоритм Бойера-Мура	7	25	9	33
Алгоритм Рабина-Карпа	~1.67	8	16	22
Вероятностный алгоритм	~1.62	7	67	72

4. ЗАКЛЮЧЕНИЕ

В результате можно сделать вывод, что каждый существующий алгоритм поиска подстроки имеет свои достоинства и недостатки, по данной причине и имеет свою наиболее подходящую прикладную область применения по сравнению с другими аналогами. Разные результаты алгоритмов в той или иной ситуации при близких асимптотических оценках временной сложности возникают вследствие опускания констант, которые могут оказать влияние на практическое время выполнения

ЛИТЕРАТУРА

- [1] Гасфилд, Д. Строки, деревья и последовательности в алгоритмах: информатика и вычислительная биология / Д. Гасфилд. – СПб.: Невский Диалект, 2003. – 654 с.
- [2] Кормен, Т.Х. Алгоритмы: построение и анализ / Т.Х. Кормен, Ч.Е. Лейзерсон, Р.Л. Ривест, К. Штайн. – М.: Вильямс, 2013. – 1328 с.
- [3] Смит, Б. Методы и алгоритмы вычислений на строках / Б. Смит. – М.: ООО «И.Д. Вильямс», 2006. – 496 с.
- [4] Rabin, M.O. Efficient randomized pattern-matching algorithms / M.O. Rabin, R.M. Karp // IBM Journal of Research and Development. – 1987. – Vol. 31(2). – P. 249-260.
- [5] Rabin, M.O. Fingerprinting by random polynomials / M.O. Rabin // Tech Report TR-CSE-03-01. Center for Research in Computing Technology, Harvard University. – 1981. – P. 1-14.
- [6] Pachocki, J. Where to use and how not to use polynomial string hashing / J. Pachocki, J. Radoszewski // Olympiads in Informatics. – 2013. – Vol. 7. – P. 90-100.

Исследование важности входных признаков при прогнозировании геомагнитного индекса алгоритмами машинного обучения

Р.Д. Владимиров
Московский государственный
университет
имени М.В. Ломоносова
Москва, Россия
vladimirov.rd16@physics.msu.ru

И.Н. Мягкова
Московский государственный
университет
имени М.В. Ломоносова
НИИ ядерной физики
имени Д.В. Скобельцына
Москва, Россия
irina@srd.sinp.msu.ru

В.Р. Широкий
Московский государственный
университет
имени М.В. Ломоносова
НИИ ядерной физики
имени Д.В. Скобельцына
Москва, Россия
shirokiy@srd.sinp.msu.ru

С.А. Доленко
Московский государственный
университет
имени М.В. Ломоносова
НИИ ядерной физики
имени Д.В. Скобельцына
Москва, Россия
dolenko@srd.sinp.msu.ru

О.Г. Баринов
Московский государственный
университет
имени М.В. Ломоносова
НИИ ядерной физики
имени Д.В. Скобельцына
Москва, Россия
obar@snp.msu.ru

Аннотация—Одним из эффективных инструментов для прогнозирования временных рядов является использование методов машинного обучения, в частности, искусственных нейронных сетей. Однако при этом необходимым этапом исследования является понижение размерности входных данных. В данной работе рассматриваются результаты понижения размерности данных на основе ранжирования входных признаков по их существенности при решении задачи прогнозирования геомагнитного индекса Dst. Для оценки относительной существенности признаков используется итеративный подход, связанный с перебором моделей-кандидатов путём отбрасывания признаков по одному.

Ключевые слова— многомерный временной ряд, прогнозирование, отбор существенных признаков, искусственные нейронные сети, магнитосфера Земли, геомагнитный индекс Dst.

1. ВВЕДЕНИЕ

Геомагнитные возмущения представляют собой один из существенных факторов космической погоды, который с развитием космической отрасли становится все более важным. Прогнозирование геомагнитных возмущений, представляет интерес, так как сильные возмущения (магнитные бури) могут стать причиной нарушений в работе телеграфных линий и радиосвязи, трубопроводов, линий электропередач и энергосетей [1]. Также они опосредованно оказывают влияние на радиационные условия в космическом пространстве, поскольку после примерно половины магнитных бурь возрастает поток релятивистских электронов внешнего радиационного пояса Земли, что сбоям в электронных микросхемах спутниковой аппаратуры.

Состояние магнитосферы Земли характеризуется геомагнитными индексами. Одним из наиболее используемых является индекс Dst. Магнитосфера Земли и представляет собой динамическую систему, будущее состояние которой зависит не только от её текущего состояния и от текущего воздействия со стороны

солнечного ветра, но и от предыстории. Это приводит к тому, что размерность используемых для прогнозирования данных, описывающих текущее состояние магнитосферы и предысторию на необходимую глубину, оказывается достаточно высока.

Одним из эффективных инструментов для прогнозирования временных рядов является использование методов машинного обучения (МО), в частности, искусственных нейронных сетей [2, 3]. Однако при этом высокая размерность входных данных может приводить к нежелательным последствиям, таким, как высокая вычислительная стоимость обучения и переучивание. Также отбор существенных входных признаков (ВП) может позволить сделать некоторые выводы о взаимосвязях различных физических величин, значения которых используются в качестве ВП.

Целью настоящей работы являлось применение адаптивного метода ранжирования ВП по их существенности на основе их отбрасывания по одному, а также оценка полученных результатов.

2. ВХОДНЫЕ ДАННЫЕ И ОПИСАНИЕ АЛГОРИТМА

В качестве входных признаков использовались исторические значения характеристик солнечного ветра (скорости SW_spd и плотности H_den) и межпланетного магнитного поля (компонент V_y и V_z в системе GSM и модуля вектора поля $|V|$), измеренные в точке Лагранжа L1 в система Солнце-Земля, и значения самого прогнозируемого индекса Dst, а также временные характеристики, описывающие фазу вращения Земли вокруг Солнца и вокруг своей оси. Каждый параметр, кроме временных характеристик, описывался 24 среднечасовыми значениями: текущим и 23 историческими значениями за последние сутки (всего 6 признаков * 24 часа + 4 = 148 ВП). Использовались данные за период с октября 1997 по конец 2016 года (тренировочный набор, из которого 20% случайно

выделялись в валидационный набор) и с начала 2017 по январь 2020 г. (тестовый набор).

Для ранжирования ВП по существенности использовался итеративный подход на основе метода МО, связанный с перебором моделей-кандидатов путём удаления ВП по одному. В данной работе для построения моделей использовалась линейная регрессия (ЛР). Общая схема работы алгоритма выглядит так:

1) Обучается модель на данных с полным набором из n ВП, и производится оценка качества модели на тестовом наборе данных.

2) Каждый ВП по очереди удаляется из набора.

3) На каждом из n уменьшенных наборов производится обучение модели, и производится оценка изменения точности по сравнению с моделью, полученной на шаге 1.

4) Тот ВП, удаление которого приводило к максимальному падению (или минимальному увеличению) ошибки прогноза по сравнению с остальными моделями, полученными на шаге 3, окончательно удаляется из набора. Ошибка прогноза полученной модели фиксируется.

5) Алгоритм повторяется с шага 1, используя полученный на шаге 4 набор в качестве полного.

6) Считается, что признаки ранжированы по важности в порядке, обратном порядку удаления: самый важный признак остается последним.

На основании зависимости среднеквадратичного отклонения прогноза модели шага 4 на тестовом наборе от количества ВП в данной модели определяется оптимальный набор признаков на основе линейной регрессии, соответствующий минимуму построенной зависимости.

Преимуществом используемого метода ЛР является его низкая вычислительная стоимость. Однако его недостатком является линейность, в то время как аппроксимируемое отображение является нелинейным. Для компенсации этого недостатка дополнительно осуществлялось построение нейросетевых моделей на отобранных наборах ВП с количеством ВП от 3 до 25. Все используемые нейронные сети представляли собой многослойные перцептроны с 1 скрытым слоем и 32 нейронами в нем. Для предотвращения переобучения использовался метод ранней остановки – обучение прекращалось спустя 500 эпох без улучшения результата на валидационном наборе. Для уменьшения влияния фактора, связанного с влиянием начальной инициализации весов на обучение, для каждого рассматриваемого случая обучалось по 5 НС, а статистические показатели их применения усреднялись.

3. РЕЗУЛЬТАТЫ

На Рис.1 показана зависимость среднеквадратичной ошибки прогнозирования на тестовом наборе данных для горизонта прогноза 1 час от номера набора (количества отброшенных ВП). Видно, что более 90% ВП могут быть исключены из рассмотрения без изменения качества прогноза. Аналогичная картина наблюдается для всех горизонтов от 1 до 6 часов.

Анализ наборов ВП, оставшихся в качестве наиболее существенных после осуществления процедуры отбора, показывает, что отобранные признаки согласуются с физическими представлениями (Рис.2).

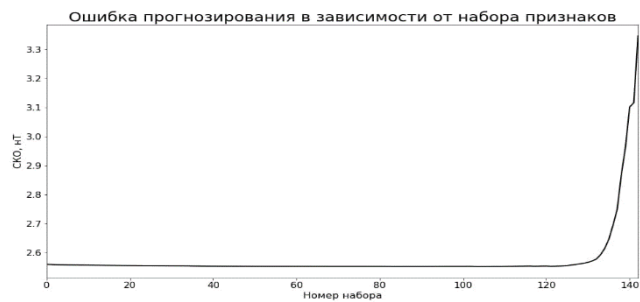


Рис. 1. Среднеквадратичная ошибка прогнозирования на тестовом наборе данных для горизонта прогноза 1 час в зависимости от номера набора (количества отброшенных ВП)

Признак	Dst	B_gsm_y					B_gsm_z					B_magn					SW_spd					H_den_SWP							
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4
0	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3	O	O	O	O	O	X	O	O	O	O	X	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
4	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
5	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
6	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
7	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
8	O	X	O	X	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
9	O	X	O	O	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O
10	X	O	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X	X	O	O	O	O	O
11	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
12	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X	X	O	O	O	O	O	O
13	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
14	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
15	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
16	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
17	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X	O	O	O	O	O
18	O	O	O	X	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
19	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X	X	O	O	O	O	O	X
20	O	O	X	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X
21	X	X	X	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X
22	X	X	X	O	O	O	O	O	O	O	X	X	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X
23	O	X	X	X	O	O	O	O	O	O	X	X	X	O	O	O	X	X	X	O	O	X	X	O	O	X	X	X	X

Рис. 2. Отобранные существенные признаки для горизонта прогноза от 1 до 6 часов. По вертикали – глубина задержки в часах

4. ВЫВОДЫ

- В результате применения ранжирования существенных входных признаков были отобраны оптимальные наборы ВП для горизонтов прогноза от 1 до 6 часов. Результаты согласуются с физическими представлениями о группах наиболее существенных ВП.

- Также удалось сократить размерность входных данных в среднем в 7.6 раз без потери качества прогнозирования. Добиться существенного улучшения качества прогнозирования путем отбора существенных ВП с помощью использованного алгоритма не удалось.

- Работы следует продолжить с использованием для построения моделей при отборе признаков нелинейных алгоритмов – градиентного бустинга или многослойных перцептронов.

- Отобранные комплекты существенных ВП могут быть зафиксированы и использоваться в дальнейшем в работах по усовершенствованию алгоритмов и методик прогнозирования.

ЛИТЕРАТУРА

[1] Лазутин, Л.Л. Мировые и полярные магнитные бури /Л.Л. Лазутин. – М.: МГУ, 2012. – 214 с.
 [2] Revallo, M. A neural network Dst index model driven by input time histories of the solar wind-magnetosphere interaction / M. Revallo, F. Valach, P. Hejda, J. Bochniček // J. Atmos. Sol. Terr. Phys. – 2014. – Vol. 110-111. – P. 9-14. DOI: 10.1016/j.jastp.2014.01.011.
 [3] Ефиторов, А.О. Прогнозирование Dst-индекса, основанное на методах машинного обучения / А.О. Ефиторов, И.Н. Мягкова, В.Р. Широкий, С.А. Доленко // Косм. иссл. – 2018. – Т. 56, № 6. – С. 353-364. DOI: 10.1134/S0010952518060035.

Разработка метода оценки качества средств технической защиты информации

Л.С. Элбакян
Волгоградский государственный университет
Волгоград, Россия
lusniak-e@yandex.ru

Л.С. Элбакян
Волгоградский государственный университет
Волгоград, Россия
liamg1829@gmail.com

Аннотация—В работе сформулированы содержательная и математическая постановки задач оценки качества технической защиты информации от утечки по техническим каналам с учетом двух параметров: показатель эффективности перекрытия каналов утечки информации (КУИ) рассматриваемыми средствами защиты информации (СЗИ) и показатель стоимости самих СЗИ. Рассматривается проблема многокритериальной оценки качества технической защиты информации с применением модели в виде двухуровневой статистической модели принятия решений в условиях неопределенности. На нижнем уровне отдельные характеристики СЗИ оцениваются на основе комбинированного критерия. Объединение значений отдельных характеристик и выбор наиболее предпочтительного производится на основе применения принципа оптимальности по Парето.

Ключевые слова— средства защиты информации, технические каналы утечки информации, многокритериальная оценка, двухуровневая модель принятия решений, принцип Паретто.

1. ВВЕДЕНИЕ

Оценка эффективности защиты информации – это непрерывный процесс, который подразумевает проведение и анализа в динамике.

На сегодняшний день проблема обеспечения технической защиты информации (ТЗИ) становится все более актуальной.

Для решения задачи оценки качества функционирования системы технической защиты информации необходимо использовать показатель качества, который позволил бы оптимизировать задачу синтеза систем защиты информации, количественно оценить эффективность функционирования системы и осуществить сравнение различных вариантов построения подобных систем.

Защита информации требует системный подход, т.е. средства и действия, используемые для обеспечения информационной безопасности, рассматриваются как единый комплекс взаимосвязанных взаимодополняющих и взаимодействующих мер.

Для успешной защиты информации пользователь должен иметь абсолютно ясное представление о возможных путях несанкционированного доступа. На практике обычно используются комбинированные способы защиты информации от утечки информации [1-2].

2. ИССЛЕДОВАТЕЛЬСКАЯ ЧАСТЬ

Рассматривается двухуровневая модель принятия решений, где предполагается наличие:

- множества технических каналов утечки информации (КУИ): $X = \{x_1, x_2, x_3, x_4\}$:
 - x_1 – акустические КУИ
 - x_2 – визуально-оптические КУИ
 - x_3 – электромагнитные КУИ
 - x_4 – материально вещественные КУИ
- множества технических СЗИ, перекрывающих эти каналы $S = \{S_1, S_2, S_3, \dots, S_{11}\}$:
 - S_1 – экранирование помещений
 - S_2 – СЗИ по цепям заземления (контроль за расположением точек заземления и длин кабеля)
 - S_3 – фильтрация опасных сигналов помехоподавляющими фильтрами
 - S_4 – пространственное зашумление
 - S_5 – линейное зашумление
 - S_6 – генераторы акустического шума
 - S_7 – система защиты переговоров
 - S_8 – оконные модуляторы
 - S_9 – СЗИ, содержащейся в отходах
 - S_{10} – средства защиты демаскирующих веществ
 - S_{11} – виброшторы

Рассмотрим пример решения задачи оценки качества защиты информации по двум критериям: по эффективности применения системы защиты информации – w_1 , по стоимости системы защиты информации – w_2 . Необходимо выбрать лучший вариант СЗИ, обеспечивающий наименьшее значения w_1 и w_2 , т.е. данные характеристики можно трактовать как функции потерь.

Необходимо выбрать лучший вариант СЗИ, обеспечивающий наименьшее значения w_1 и w_2 , т.е. данные характеристики можно трактовать как функции потерь. Значения функции потерь $U_1 = \|u_1(x_i, S_j)\|$ характеристики эффективности СЗИ для различных видов ТКUI и значения функции потерь $U_2 = \|u_2(x_i, S_j)\|$ стоимости данных СЗИ оценены экспертами. Значения функций потерь U_1 и U_2 характеристик w_1 и w_2 оцениваются в разных системах измерений: эффективность перекрытия – в процентах, стоимость СЗИ – в рублях. При этом СЗИ, имеющие 100% оценочные значения, свидетельствовало о том, что данный способ не может быть применен для защиты выбранных КУИ. Для удобства дальнейшего сравнения значения функций потерь необходимо нормализовать. Значения функций потерь в шкале от 0 до 10 баллов представлены так, чтобы лучшие значения имели меньшее число баллов, а худшие – большее.

Задача оценки качества СЗИ состоит в выборе наилучшего варианта с помощью решения двухуровневой задачи оптимизации [3-5]. На верхнем уровне решается задача оценки функции потерь U_k . На нижнем уровне для $k = 1$ и 2 для каждого $x_i \in X$ при параметрах $\lambda_1, \lambda_2 \in [0,1]$

(определяются с учетом априорной информированности лиц, принимающих решение) определяются функции $y_1(x_i)$ и $y_2(x_i)$ оценки качества СЗИ для соответствующей функции потерь U_1 и U_2 .

3. ОЦЕНКА КАЧЕСТВА СЗИ С ПРИМЕНЕНИЕМ ПРИНЦИПА ОПТИМАЛЬНОСТИ ПАРЕТТО

На верхнем уровне решается задача оценки функции потерь U_k с применением принципа оптимальности Парето [6-7].

На основании данных построены диаграмма Парето эффективности и диаграмма Парето стоимости СЗИ (рис.1).

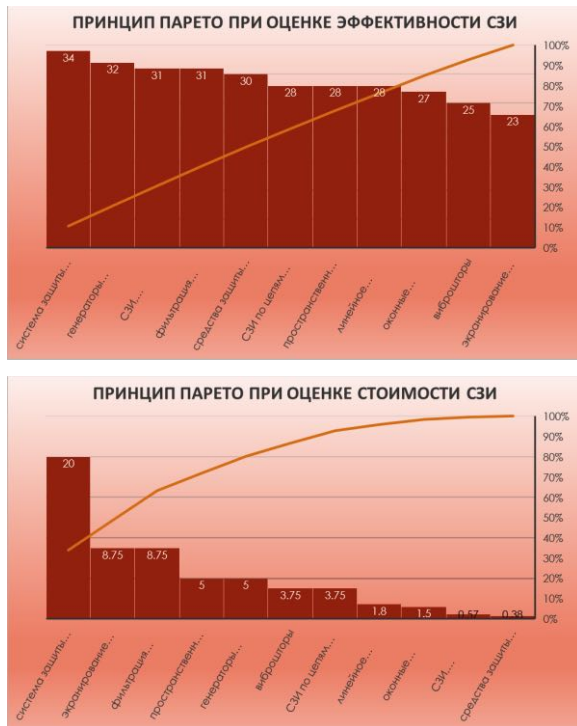


Рис.1. Принцип Парето при оценке эффективности и стоимости СЗИ

4. ОЦЕНКА КАЧЕСТВА СЗИ С УЧЕТОМ АПРИОРНОЙ ИНФОРМИРОВАННОСТИ ЛИЦ, ПРИНИМАЮЩИХ РЕШЕНИЕ

На нижнем уровне для $k=1$ и 2 для каждого $x_i \in X$ при параметрах $\lambda_1, \lambda_2 \in [0,1]$ (определяются с учетом априорной информированности лиц, принимающих решение) определяются функции $y_1(x_i)$ и $y_2(x_i)$ оценки качества СЗИ для соответствующей функции потерь U_1 и U_2 :

Далее оценка качества СЗИ полностью ориентирован на выбор параметра λ_1 , который зависит от лиц, принимающих решение и может принимать значения от 0 до 1.

Таким образом, получена итоговая оценка качества технических СЗИ в зависимости от ТКУИ по

эффективности $y_1(x_i, \lambda_1)$ и стоимости $y_2(x_i, \lambda_1)$ в виде Евклидово расстояния.

На основании полученных математических расчетов, можно предположить, что при выбранные СЗИ могут использоваться в качестве эффективных средств следующих ТКУИ:

- акустические КУИ
- электромагнитные КУИ
- материально вещественные КУИ

Причем наиболее предпочтительны в случае защиты от акустических каналов утечки информации.

5. ВЫВОДЫ

В работе разработана математическая модель оценки качества средств технической защиты информации: метод математического анализа качества систем технической защиты информации с применением двухуровневой модели принятия решений в условиях неопределенности. Сформулированы содержательная и математическая постановки задач оценки качества технической защиты информации от утечки по техническим каналам с учетом двух параметров: показатель эффективности перекрытия КУИ рассматриваемыми СЗИ и показатель стоимости самих СЗИ. Проведены экспериментальные исследования оценки с применением двухуровневой оценки: на верхнем уровне проблема решается с применением принципа оптимальности Парето, на нижнем уровне - с применением математических расчетов, которые полностью зависят от выбора лиц, принимающих решение.

ЛИТЕРАТУРА

- [1] Шакиров, В.А. Многокритериальная оценка альтернатив на основе теории полезности в условиях неопределенности предпочтений лица, принимающего решения / В.А. Шакиров // Нечеткие системы и мягкие вычисления. – 2018. – Т. 13, № 1. – С. 17-35.
- [2] Рыков, А.С. Методы системного анализа: Многокритериальная и нечеткая оптимизация, моделирование и экспертные оценки / А.С. Рыков. – М.: Экономика, 1999. – 192 с.
- [3] Макаров, О.Ю. Методы многокритериальной оценки / О.Ю. Макаров, В.В. Цветков // Вестник Воронежского государственного технического университета. – 2009. – Т. 5, № 11. – С. 133-135.
- [4] Ларионова, С.Л. Метод принятия решений в условиях неопределенности для обеспечения информационной безопасности / С.Л. Ларионова, Л.И. Товпеко // Инновации и инвестиции. – 2020. – Т. 6. – С. 114-118.
- [5] Рыков, А.А. Модель оценки отдельных характеристик качества информационных систем в условиях неопределенности / А.А. Рыков // 3 междунар. конф. «Идентификация систем и задачи управления». – Москва: Ин-т проблем управления, 2004.
- [6] Ногин, В.Д. Обобщенный принцип Эджворта-Парето и границы его применимости. Экономика и методические методы / В.Д. Ногин // Журнал вычислительной математики и математической физики. – 2005. – Т. 41, № 3. – С. 128-134.
- [7] Арутюнов, В.В. Об интерпретации закона Парето в сфере информационной безопасности / В.В. Арутюнов // Вестник Московского финансово-юридического университета. – 2015. – С. 53-62.

Определение вероятности потери информации на частично записанном оптическом диске однократной записи при длительном хранении в электронном архиве

А.В. Чернышов
Мытищинский филиал МГТУ им. Н. Э. Баумана
Мытищи, Россия
sch-ru@yandex.ru

Аннотация—Рассмотрена задача определения вероятности потери информации на частично записанном оптическом диске однократной записи при длительном хранении в электронном архиве. Предложено два подхода к решению задачи. Приведены результаты численных расчётов, показавшие хорошее согласование искомого параметра, вычисленного обоими методами.

Ключевые слова— электронный архив длительного хранения информации, оптический диск однократной записи, вероятность потери информации.

1. ВВЕДЕНИЕ

Несмотря на развитие облачных технологий хранения больших объёмов информации, задача архивного (длительного) хранения электронных документов организаций должна по-прежнему решаться путём создания локальных электронных архивных хранилищ этих организаций. Международный опыт [1-4] склоняется к необходимости использования носителей однократной записи (WORM), к которым сегодня можно отнести только оптические диски типов CD-R, DVD+-R, BD-R.

При проектировании, создании и последующей эксплуатации такого архива естественным требованием является полная сохранность архивной информации. Для дисков однократной записи информация может быть потеряна частично или полностью из-за старения самого носителя — деградации диска [5, 6].

Проблема деградации оптических дисков с записью в процессе хранения мало изучена. Современные производители оптических дисков однократной записи предпочитают ограничиваться ускоренными тестами своих изделий [7], исходя из результатов которых обещают потребителям так называемый «срок службы» диска. Сегодня «срок службы» обычно называется в пределах 30 лет, иногда чуть больше.

Проблема заключается в том, что в течение указанного «срока службы» из-за деградации дисков всё равно существует опасность потери части хранимой информации из-за того, что какие-то диски из хранимого массива перестанут читаться частично или полностью. Она может быть решена путём создания архива с несколькими копиями хранимой информации и регулярным тестированием целостности информации на дисках [8, 9]. В случае обнаружения деградировавших дисков они могут быть заменены в архиве на новые с записью на них информации из запасных копий.

Однако для того, чтобы определить параметры эксплуатации архива (необходимое количество запасных копий, время, затрачиваемое на тестирование и замену дисков, необходимое количество ежегодно приобретаемых новых дисков для замены деградировавших) необходимо знать значение вероятности потери информации на одном оптическом диске q . А значение этого параметра никто из производителей дисков не сообщает.

На основании анализа доступных немногочисленных источников [6, 10, 11], а также по результатам проведённых собственных натуральных экспериментов [12, 13] автор может утверждать, что для дисков с заполнением, близким к их ёмкости, для практических расчётов надёжности архива можно принять вероятность потери информации на одном диске $q=0,1$.

Заполнение каждого диска близко к его максимальной ёмкости для его размещения в архиве является логичным и экономически оправданным.

Однако на практике встречаются специализированные архивы, при создании которых к заполнению дисков могут выдвигаться особые требования. Например, требование размещать на одном диске не больше одного файла или только файлы, относящиеся к какой-либо логически связанной группе.

В этом случае на диске будет оставаться незанятое пространство, деградация которого, не затрагивающая часть диска с записью, никак не скажется на воспроизведении с него информации. Следовательно, вероятность потери информации (обозначим её q^*) будет ниже.

На сегодняшний день отсутствуют данные, позволяющие определить вероятность потери информации в процессе хранения на не полностью заполненных дисках, которые были бы получены в результате натуральных экспериментов. Однако для практических расчётов таких архивов знание значения параметра необходимо. Попытаемся его оценить.

2. ОЦЕНКА ВЕРОЯТНОСТИ ПОТЕРИ ИНФОРМАЦИИ

А. Оценка геометрическим методом

Одним из вариантов оценки является геометрический подход. В этом случае будем просто считать, что вероятность потери информации линейно уменьшается с уменьшением занятой информацией площади на диске. Причём занятая площадь линейно зависит от объёма

записанной информации, выраженного в процентах ёмкости диска. То есть, например, для диска BD-R (полная ёмкость 25 гигабайт) при записи 5 гигабайт (20 % от полной ёмкости диска) получим $q^*=0,02$.

Б. Оценка вероятностным методом

Альтернативным подходом является попытка вычислить значение методами теории вероятностей.

Пусть архив создан из дисков ёмкостью V и q — вероятность потери хотя бы какой-то части информации на этом диске при хранении в течение года.

Пусть v — размер одного блока на диске и p — вероятность потери информации на нём. В данном случае не принципиально, о каком конкретно блоке в структуре записи информации на диске идёт речь. Но для корректности рассуждений будем считать, что запись информации на диск может выполняться только кратно размеру блока, то есть искомая величина соответствует вероятности потери информации в этом блоке.

Тогда всего на диске блоков $n=V/v$. И можно записать

$$q = \sum_{i=1, n} p^i (1-p)^{n-i} C_n^i \quad (1)$$

Обратим внимание, что выражение

$$\sum_{i=0, n} p^i (1-p)^{n-i} C_n^i = (p+(1-p))^n \quad (2)$$

представляет собой бином Ньютона. Следовательно

$$q = (p+(1-p))^n - C_n^0 (1-p)^n p^0 \quad (3)$$

В результате получим

$$p = 1 - (1-q)^{1/n} \quad (4)$$

Если теперь задаться средней долей заполненности каждого диска a ($a < 1$), то можно записать, что вероятность потери информации на диске q^* можно вычислить как

$$q^* = 1 - (1-p)^{an} \quad (5)$$

3. СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Сравним результаты обоих методов расчёта. Для этого построим несложную таблицу в LibreOffice Calc для значений от 0,1 до 1,0 с шагом 0,1. Будем выполнять вычисления для дисков типа BD-R ($V=25$ гигабайт). Примем размер одного блока $v=2048$ байт. Результаты расчётов приведены в таблице 1.

В приведённой таблице числа сознательно даны именно в той форме, как показывает их LibreOffice Calc. Можно видеть, что при небольших значениях вероятностный метод даёт значения q^* , незначительно превышающие геометрический метод. Выполнив округление до сотых стандартным методом, принятым в математике, получим совпадение значений в обоих методах.

4. ЗАКЛЮЧЕНИЕ

Учитывать ли получившиеся расхождения при проектных расчётах архива, должен решить сам проектировщик архива. К тому же для полностью заполненного диска результат вычисления для вероятностного метода явно содержит погрешности вычислений, что заставляет подумать о том же и для других результатов.

В целом можно констатировать, что оба метода вычисления значений q^* не противоречат друг другу, а скорее подтверждают результаты друг друга.

ТАБЛИЦА 1. СРАВНЕНИЕ РЕЗУЛЬТАТОВ ВЫЧИСЛЕНИЙ

a	q^*	
	Геометрический метод	Вероятностный метод
0,1	0,01	0,0104807417
0,2	0,02	0,0208516375
0,3	0,03	0,0311138387
0,4	0,04	0,0412684843
0,5	0,05	0,0513167017
0,6	0,06	0,0612596064
0,7	0,07	0,071098302
0,8	0,08	0,0808338808
0,9	0,09	0,0904674235
1,0	0,10	0,0999999995

ЛИТЕРАТУРА

- [1] ISO/TR 18492:2005 Long-term preservation of electronic document-based information / Last reviewed and confirmed in 2013. – URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=38716 (26.07.2021).
- [2] Zheng, J. NIST/Library of Congress Optical Disc Longevity Study: Final Report / J. Zheng, O.T. Slattery. – 2007. – 32 p. URL: https://www.loc.gov/preservation/resources/rt/NIST_LC_OpticalDiscLongevity.pdf (11.05.2016).
- [3] Haus, G. Preserving cultural heritage: A new approach to increase the life expectancy of optical discs / G. Haus, C. Polizzi, A. Visconti // Journal of Cultural Heritage. – 2018. – Vol. 29. – P. 67-74. DOI: 10.1016/j.culher.2017.08.004.
- [4] Bradley, K. Risks Associated With the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections / K. Bradley // Strategies and Alternatives. – UNESCO, Paris, 2006. – 31 p.
- [5] Smith, E. When Discs Die / E. Smith // Tedium // URL: <https://tedium.co/2017/02/02/disc-rot-phenomenon/> (18.11.2017).
- [6] Пилипчук, М.И. Рекомендации по обеспечению сохранности информации, записанной на оптических дисках (Тестирование выборочного массива документов федеральных архивов) / М.И. Пилипчук, А.Н. Балакирев, Л.В. Дмитриева, Г.З. Залаев. – М.: РГАНТД, 2011. – 52 с.
- [7] ISO/IEC 10995:2011(E) Information technology - Digitally recorded media for information interchange and storage // Test method for the estimation of the archival lifetime of optical media. – 2015. – 28 p.
- [8] Zalaev, G.Z. Some Issues of Long-Term Storage of Electronic Documents / G.Z. Zalaev, N.E. Kalenov, V.A. Tsvetkova // Sci. Tech. Inf. Proc. – 2016. – Vol. 43(4). – P. 268-274. DOI: 10.3103/S0147688216040110.
- [9] Corrado, E.M. Digital Preservation for Libraries, Archives, and Museums / E.M. Corrado, H.L. Rowman. – Moulaison & Littlefield Publishers, 2014. – 294 p.
- [10] van Ooijen, P.M.A. Accessibility of Data Backup on CD-R after 8 to 11 years / P.M.A. van Ooijen, A.R. Viddeleer, F. Meijer, M. Oudkerk // J Digit Imaging. – 2010. – Vol. 23. – P. 95. DOI: 10.1007/s10278-008-9161-9.
- [11] Устинов, В. Хранение данных на CD- и DVD-дисках: на наш век хватит? / В. Устинов // BROADCASTING: Телевидение и радиовещание. – 2006. – № 4. URL: http://www.broadcasting.ru/articles2/Oborandteh/hranenie_dannyh_na_CD_DVD_diskah (10.12.2018).
- [12] Чернышов, А.В. Оценка вероятности выхода из строя оптического диска DVD-R с записью по результатам натурального эксперимента / А.В. Чернышов // Современные технологии в задачах управления, автоматизации и обработки информации: Сборник трудов XXVII Международной научно-технической конференции. – Тамбов: ТГТУ, 2018. – С. 272.
- [13] Чернышов, А.В. Определение значения вероятности выхода из строя однократно записываемого оптического диска с записью в процессе хранения в долговременном архиве / А.В. Чернышов // Лесной вестник. – 2019. – Т. 23, № 4. – С. 82-87. DOI: 10.18698/2542-1468-2019-4-82-87.

Применение цифровых двойников для анализа прочности костей на основе данных компьютерной томографии

О.В. Герасимов
Казанский (Приволжский) федеральный университет
Казань, Россия
valeriy.karasikov@gmail.com

К.Р. Шарафутдинова
Казанский (Приволжский) федеральный университет
Казань, Россия
KaRSharafutdinova@stud.kpfu.ru

Р.Р. Рахматулин
Казанский (Приволжский) федеральный университет
Казань, Россия
RRRakhmatulin@stud.kpfu.ru

Т.В. Балтина
Казанский (Приволжский) федеральный университет
Казань, Россия
tvbaltina@kpfu.ru

М.Э. Балтин
Казанский (Приволжский) федеральный университет
Казань, Россия
baban.bog@mail.ru

А.О. Федянин
Казанский (Приволжский) федеральный университет
Казань, Россия
artishock23@gmail.com

Аннотация—В данной работе рассматривается способ обработки данных компьютерной томографии (КТ) и метод последующего расчёта и анализа результатов на основе построения конечно-элементного (КЭ) ансамбля. Цель исследования заключается в описании общего алгоритма построения численной модели для статического расчёта пористых объектов на основе данных КТ. Количественная оценка значений основывалась на нормированной ошибке энергии. В качестве модельной задачи рассматривался трёхточечный изгиб бедренной кости свиньи. Результаты численного эксперимента сравнивались с данными натурального испытания.

Ключевые слова— цифровой прототип, численное моделирование, костные органы, конечно-элементный анализ, негомогенные среды.

1. ВВЕДЕНИЕ

Применение данных КТ при построении численных моделей выступает основным подходом к созданию цифрового прототипа элементов негомогенных сред [1]. Подобные задачи особенно актуальны в клинической практике [2]. Существует несколько подходов, основанных на применении данных с изображений. В основе первого лежит построение МП-распределения [3]. Другой предполагает сведение анизотропии к ортотропии путём вычисления констант из численных экспериментов [4]. В данной работе предлагается к рассмотрению метод прямого учёта свойств материала с помощью взвешенного интегрирования локальной матрицы жёсткости (ЛМЖ). Цель исследования заключается в реализации методики статического расчёта негомогенных пористых элементов на основе трёхмерного изопараметрического КЭ сплошной среды с линейной аппроксимацией, построенного по данным КТ.

2. МАТЕРИАЛЫ И МЕТОДЫ

В основе подхода лежало интегрирование ЛМЖ с весовой функцией, соответствующей данным КТ [5]. Бинаризируя исходный массив, производилась гетерогенизация материала. В качестве метода интегрирования был выбран метод средних прямоугольников [6]. Напряжённо-деформированное состояние (НДС) локально усреднялось по объёму [7]. Полученная из расчётов величина ошибки энергии по

напряжениям нормировалась по энергии деформации [8]. Построение расчётной области осуществлялось путём удаления элементов с низким содержанием кости из исходной регулярной сетки. Эксперименты проводились на плечевой кости Вьетнамской свиньи. Протокол эксперимента был одобрен Комитетом по уходу за животными Казанского государственного медицинского университета (протокол №5 от 20 мая 2020 года) [9]. Механические параметры соответствовали костной ткани. Перемещения прикладывались к диафизу, закрепление дистальных участков соответствовало жёсткому закреплению. Натурный эксперимент проводился в условиях действия поперечной нагрузки. Сравнение результатов осуществлялось на основе вычисления эквивалентной силы.

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Расчёт производился на двух КЭ-сетках различной плотности. Особое внимание уделялось областям максимальных напряжений по Мизесу. Результаты показали, что максимальные значения ошибки энергии возникают на границах с наименьшим относительным содержанием костной ткани. Значения в интересующих областях составили менее 25%, кроме зон, подверженных кинематическим граничным условиям (~ 50%). Минимальная ошибка энергии (20%) и максимальные напряжения по Мизесу соответствовали области приложения нагрузки. В регионе максимальных напряжений по Мизесу первая главная компонента достигает максимума, а третья – минимума, что объясняет возникновение трещины при проведении натурального эксперимента, так как значения превышают предел прочности. Относительная ошибка в сравнении с данными натурального испытания составила от 3 до 15%.

4. ЗАКЛЮЧЕНИЕ

Реализован метод обработки данных КТ на основе распределения плотности материала в объёме исследуемого образца. Осуществлён КЭ-анализ НДС бедренной кости свиньи. Проведена оценка полученных результатов на основе значений ошибки энергии. Поставлен натурный эксперимент в условиях действия трёхточечного изгиба.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке стипендии Президента Российской Федерации молодым учёным и аспирантам (№ стипендии СП-4182.2022.4).

ЛИТЕРАТУРА

- [1] Marwa, F. Improved USCT of Paired Bones Using Wavelet-based Image Processing / F. Marwa, E.Y. Wajih, L. Philippe, M. Mohsen // IJGSP. – 2018. – Vol. 10(9). – P. 1-9. DOI: 10.5815/ijgsp.2018.09.01.
- [2] Kichenko, A.A. Experimental determination of the fabric tensor for cancellous bone tissue / A.A. Kichenko, V.M. Tverier, Y.I. Nyashin, A.A. Zaborskikh // Russ. J. Biomech. – 2011. – Vol. 15(4). – P. 66-81.
- [3] Sachenkov, O. Determination of muscle effort at the proximal femur rotation osteotomy / O. Sachenkov, R. Hasanov, P. Andreev, Y. Konoplev // IOP Conf. Ser. Mater. Sci. Eng. – 2016. – Vol. 158(1). – P. 012079.
- [4] Ridwan-Pramana, A. Finite element analysis of 6 large PMMA skull reconstructions: A multi-criteria evaluation approach / A. Ridwan-Pramana, P. Marcian, L. Borak, N. Narra, T. Forouzanfar, J. Wolff // PLoS ONE. – 2017. – Vol. 12. – P. e0179325. DOI: 10.1371/journal.pone.0179325.
- [5] Relations of mechanical properties to density and CT numbers in human bone / J.Y. Rho, M.C. Hobatho, R.B. Ashman // Medical Engineering & Physics. – 1995. – Vol. 17(5). – P. 347-355.
- [6] Воробьёв, О.В. Конечно-элементная оценка деформированного состояния по данным компьютерной томографии / О.В. Воробьёв, Е.В. Семёнова, Д.А. Мухин, Е.О. Стаценко, Т.В. Балтина, О.В. Герасимов // Вестник ПНИПУ. Мех. – 2021. – № 2. – С. 44-54. DOI: 10.15593/perm.mech/2021.2.05.
- [7] Giovannelli, L. Direct medical image-based Finite Element modelling for patient-specific simulation of future implants / L. Giovannelli, J.J. Rodenas, J.M. Navarro-Jimenez, M. Tur // Finite Elem. Anal. Des. – 2017. – Vol. 136. – P. 37-47. DOI: 10.1016/j.finel.2017.07.010.
- [8] Gerasimov, O.V. Bone Stress-Strain State Evaluation Using CT Based FEM / O.V. Gerasimov, N.V. Kharin, A.O. Fedyanin, P.V. Bolshakov, M.E. Baltin, E.O. Statsenko, F.O. Fadeev, R.R. Islamov, T.V. Baltina, O.A. Sachenkov // Front. Mech. Eng. – 2021. – Vol. 7. – P. 688474. DOI: 10.3389/fmech.2021.688474.
- [9] Fadeev, F. Combined Supra- and Sub-Lesional Epidural Electrical Stimulation for Restoration of the Motor Functions after Spinal Cord Injury in Mini Pigs / F. Fadeev, A. Eremeev, F. Bashirov, R. Shevchenko, A. Izmailov, V. Markosyan, M. Sokolov, J. Kalistratova, A. Khalitova, R. Garifulin, R. Islamov, I. Lavrov // Brain Sciences. – 2020. – Vol. 10. – P. 744. DOI: 10.3390/brainsci10100744.

Повышение достоверности обнаружения состояний гонки данных многопоточных системах

А.Ю. Лобачев

Самарский государственный университет путей сообщения
Самара, Россия
lobachevandreii1996@gmail.com

В.А. Засов

Самарский государственный университет путей сообщения
Самара, Россия
vzasov@mail.ru

Аннотация—Предложена методика коллективного тестирования группой детекторов гонки произвольных программ для обнаружения состояний гонки данных многопоточных системах. Разработан алгоритм для генерации программ для тестирования детекторов гонки данных. Предложенные методика и алгоритм позволяют повысить достоверность обнаружения состояний гонки данных в многопоточных системах.

Ключевые слова— системы, многопоточные, гонки, данные, обнаружение, алгоритмы, достоверность.

1. ВВЕДЕНИЕ

Состояния гонки данных возникают в многопоточных параллельных системах, когда несколько потоков одновременно обращаются к одним тем же данным и хотя бы один из потоков выполняет операцию записи, причем порядок этих обращений точно не определен [1].

В этих ситуациях результаты работы программы не детерминированы и зависят от времени или порядка исполнения потоков [1, 2]. Проблема гонки данных приводит к ошибкам, которые при попытке обнаружения классическими методиками отладки могут никак не проявлять себя и оставаться скрытыми. Это может оказывать серьезное негативное воздействие на надежность программного обеспечения.

Алгоритмы и реализующие эти алгоритмы программы обнаружения состояний гонки данных, в дальнейшем детекторы гонки данных, разделяются на статические и динамические.

Статические детекторы гонки анализируют исходный код программы и строят её модель, основывающуюся, как правило, на графе потока выполнения и графе использования объектов синхронизации [1, 2, 3].

Динамические детекторы гонки анализируют программу во время её работы, собирая информацию при трассировке программы об истории обращений к памяти и выполнении операций синхронизации [4].

В настоящее время обобщенный универсальный алгоритм обнаружения гонки данных для произвольных программ не создан. Разработанные детекторы гонки являются специализированными и используют оригинальные методы, основанные на особенностях реализуемых в многопоточных системах вычислительных процессов, применяемых структур данных и др.

Например, алгоритм, ESC/Java использует расширенный подход статической проверки ограничений времени компиляции программного кода основанный на

автоматическом доказательстве теорем [5]. Алгоритм Chord основан на методике многофазного потокозависимого анализа состояний гонки, который в своей основе применяет языковые средства рефлексии для анализа работы потоков и состояния данных [6]. Алгоритм RccJava основан на расширении системы типов Java для большого числа популярных паттернов проектирования многопоточного кода, что позволяет обнаруживать несинхронизированные операции над данными [7]. Алгоритм RecPlay собирает трассу выполнения потоков в рантайме, а затем на основании собранных данных проверяет отношения happens-before на предмет возникновения гонки используя механизм векторных часов [6].

При индивидуальном использовании детекторов их специализация, а также стохастическая природа возникновений гонки данных в произвольных программах являются причинами, снижающими достоверность результатов анализа о количестве и местах обнаружения гонки данных в программах. Технология применения разработанных детекторов недостаточно автоматизирована и ориентирована на квалифицированных специалистов, что усложняет этот вид тестирования программного обеспечения.

Целью работы является разработка методики и инструментов для ее реализации, позволяющие повысить достоверность обнаружения состояний гонки данных в многопоточных системах.

2. КОЛЛЕКТИВНОЕ ТЕСТИРОВАНИЕ ДЛЯ ОБНАРУЖЕНИЯ СОСТОЯНИЙ ГОНКИ ДАННЫХ В МНОГОПОТОЧНЫХ СИСТЕМАХ

Для повышения достоверности обнаружения состояний гонки данных в произвольных программах предлагается производить тестирование группой детекторов и результаты тестирования определять на основе интегрированной обработки результатов анализа полученных группой детекторов. Реализацию коллективного тестирования предлагается производить на разработанном автоматизированном программном комплексе.

Программный комплекс позволяет вычислять основные показатели, определяющие полноту и точность информации о состояниях гонки. Это общее количество обнаруженных состояний гонки, количество ложных состояний гонки, количество пропущенных состояний гонки и соотношения этих показателей.

Пример результатов коллективного тестирования клиент-серверного приложения, представляющего

многопоточный сервис формирования xml документов и содержащего примерно 20000 строк кода, приведен в таблице 1.

ТАБЛИЦА 1. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ КЛИЕНТ-СЕРВЕРНОГО ПРИЛОЖЕНИЯ

Применяемые алгоритмы	Обнаруженные гонки	Ложные гонки	Время выполнения (с)
ESC/Java	24	11	442
Chord	23	7	384
RccJava	27	9	470
RecPlay	11	0	301(с накладными расходами 342)

На рис. 1 приведены результаты тестирования вышеуказанного приложения, показывающие количество детекторов (всего 4) обнаруживших состояния гонки на каждом из 20 анализируемых участков программы. Если под ложным состоянием гонки считать состояния гонки обнаруженное только одним детектором из группы, то количество ложных состояний гонки 5 (4, 5, 10, 13 и 19 участки программы).

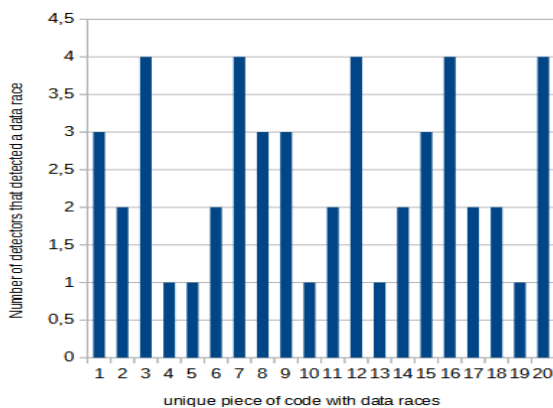


Рис. 1. Результаты тестирования приложения: количество детекторов, обнаруживших гонки данных на каждом из участков кода программы

Устранение состояний гонки данных осуществляется применением средств синхронизации потоков (семафоров, мониторов и др.), что снижает быстродействие программ из-за уменьшения степени параллелизма выполнения потоков. Поэтому игнорирование ложных состояний гонки позволяет избегать ненужную синхронизацию и, следовательно, не снижать быстродействие программ.

Для определения количества пропущенных состояний гонки данных в работе предлагается алгоритм псевдослучайной генерации тестовых программ с заранее известным наличием или отсутствием секций с конкурентным изменением данных. Программный комплекс позволяет задавать размер тестовых программ, количество критических секций в отдельной сгенерированной программе. После завершения тестирования детекторов гонки можно определить количество обнаруженных состояний гонки, соотношения обнаруженных, пропущенных и ложных состояний гонки. Результаты тестирования можно определять по всему набору программ и детально.

На рис. 2 приведены результаты тестирования вышеуказанных 4 детекторов гонки на тестовой

программе состоящей примерно из 8500 строк кода и содержащей 37 участков кода с возможными состояниями гонки данных. Тестирование показало 5 пропущенных группой детекторов состояний гонки.

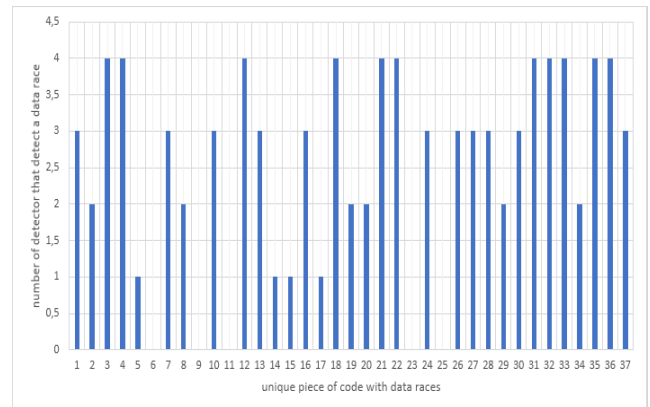


Рис. 2. Результаты тестирования группы детекторов гонки с целью определения количества пропущенных состояний гонки

3. ЗАКЛЮЧЕНИЕ

Предложена методика и программный комплекс для коллективного тестирования группой детекторов гонки произвольных программ для обнаружения состояний гонки данных многопоточных системах. Разработан алгоритм для генерации программ для тестирования детекторов гонки данных.

Предложенные методика и алгоритм позволяют повысить достоверность обнаружения состояний гонки данных в многопоточных системах и надежность разрабатываемого программного обеспечения.

ЛИТЕРАТУРА

- [1] Битнер, В.А. Построение универсального линейризованного графа потока управления для использования в статическом анализе кода алгоритмов / В.А. Битнер, Н.В. Заборовский // Модел. и анализ информ. систем. – 2013. – Т. 20, № 2. – С. 166-177.
- [2] Barabanova, P. Modeling and Investigating a Race Condition Detection Algorithm for Multithread Computational Systems / P. Barabanova, V. Zasov // XXI International Conference Complex Systems: Control and Modeling Problems (CSCMP). – Samara, 2019. – P. 356-359. DOI: 10.1109/CSCMP45713.2019.8976855.
- [3] Барабанова, П.С. Моделирование и исследование алгоритма обнаружения состояний гонки в многопоточных вычислительных системах / П.С. Барабанова, В.А. Засов // Труды XXI международной конференции «Проблемы управления и моделирования в сложных системах». – Самара: ООО «Офорт», 2019. – С. 195-198.
- [4] Трифанов, В.Ю. Динамические средства обнаружения гонок в параллельных программах / В.Ю. Трифанов, Д.И. Цителов // Компьютерные инструменты в образовании. – 2011. – № 5. – С. 3-15.
- [5] Flanagan, C. Extended static checking for Java / C. Flanagan, K.R.M. Leino, M. Lillibridge, G. Nelson, J.B. Saxe, R. Stata // Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation. – 2002. – Vol. 37(5). – P. 234-245. DOI: 10.1145/512529.512558.
- [6] Java Programm Representation [Electronic resource]. – Access mode: https://www.seas.upenn.edu/~mhnai/chorde/user_guide/program.html
- [7] Flanagan, C. Type-Based Race Detection for Java / C. Flanagan, S.N. Freund // Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation. – 2000. – Vol. 35(5). – P. 219-232. DOI:10.1145/349299.349328.

Анализ методов отбора признаков для задачи классификации когнитивных состояний и состояний покоя по данным ЭЭГ

М.Ш. Муртазина

Новосибирский государственный технический университет
Новосибирск, Россия
murtazina@corp.nstu.ru

Т.В. Авдеевко

Новосибирский государственный технический университет
Новосибирск, Россия
avdeenko@corp.nstu.ru

Аннотация—В работе исследуются вопросы применения методов отбора признаков, входящих в пакет программ Weka, для решения задачи классификации когнитивных состояний и состояний покоя по данным ЭЭГ. Для проведения эксперимента использован набор данных «EEG During Mental Arithmetic Tasks». Из записей ЭЭГ для 19 скальповых электродов извлечены параметры Хьорта и мощность в дельта-, тета-, альфа- и бета-частотных диапазонах. На первом этапе эксперимента на полученном признаковом пространстве выполнено обучение классифицирующих моделей, на втором – выполнено исследование изменения точности классификаторов, обученных на том же наборе алгоритмов на подмножестве признаков, полученных при помощи методов отбора признаков. Для получения подмножества признаков использованы оценщики атрибутов CfsSubsetEval и WrapperSubsetEval.

Ключевые слова— отбор признаков, классификация, состояние покоя, когнитивные состояния, ЭЭГ, Weka.

1. ВВЕДЕНИЕ

В последние годы все больше исследований посвящено вопросам применения данных электроэнцефалографии (ЭЭГ) для классификации когнитивных состояний и состояний покоя на основе признаков ЭЭГ [1,2,3]. Отбор признаков является важным шагом в классификации умственных состояний по данным ЭЭГ. На практике наиболее значимые для задачи классификации умственных состояний по данным ЭЭГ признаки априори неизвестны. По этой причине в признаковое пространство может быть включено большое количество признаков, извлеченных из сигналов ЭЭГ от каждого скальпового электрода. При этом количество электродов определяется целями исследования и используемым оборудованием.

Методы отбора признаков эффективно используются при подготовке данных для интеллектуального анализа и машинного обучения. Целью процедуры отбора признаков является поиск подмножества признаков, которые являются значимыми для прогноза. Отбор признаков считается успешным, если размерность признакового пространства уменьшена, а точность классификатора стала выше или не изменилась.

Цель данного исследования – провести анализ методов отбора признаков, предоставляемых пакетом программ Weka, которые могут быть использованы для уменьшения признакового пространства ЭЭГ высокой размерности.

Работа организована следующим образом. В разделе 1 обосновывается актуальность темы исследования. В

разделе 2 дается описание материалов и методов исследования. В разделе 3 представлены результаты экспериментальной части исследования. В разделе 4 подводятся итоги работы.

2. МАТЕРИАЛЫ И МЕТОДЫ

А. Формирование признаков

Для проведения эксперимента использован набор данных «EEG During Mental Arithmetic Tasks», который включает записи ЭЭГ в состоянии покоя (три минуты) и выполнения арифметических вычислений (одна минута) для 36 субъектов [4]. Файлы с записями электроэнцефалограмм представлены в наборе данных в формате EDF.

Для извлечения признаков использованы данные 19 скальповых электродов в соответствии с международной схемой 10-20. Из записей ЭЭГ для каждого электрода извлечены параметры Хьорта (активность, мобильность и сложность) и мощность в дельта-, тета-, альфа- и бета-частотных диапазонах. Оценка спектральной плотности мощности (PSD) произведена методом Уелча. Размер окна для извлечения признаков составил одну минуту. Из исходного набора были использованы записи ЭЭГ для 34 субъектов, поскольку длины записей ЭЭГ в состоянии покоя у двух испытуемых были менее трех минут. Для каждого субъекта было получено по четыре экземпляра данных. Три из них соответствуют состоянию покоя с закрытыми глазами, один – состоянию, связанному с когнитивной нагрузкой, обусловленной выполнением арифметических вычислений.

Для извлечения признаков из записей ЭЭГ использована Python-библиотека eeglib, которая предоставляет инструменты, позволяющие пользователю легко создать целый набор данных, совместимых с большинством Python-библиотек анализа данных. Эта возможность обеспечивается созданием объекта DataFrame с извлекаемыми признаками [5]. Для автоматизации процесса формирования набора данных в формате ARFF, используемом пакетом программ Weka, был создан Python-скрипт.

Б. Методы отбора признаков в Weka

Для отбора признаков и обучения классификаторов использовано свободное программное обеспечение для анализа данных и машинного обучения Weka. Данный программный продукт может работать на любой платформе, поскольку является java-приложением.

Процедура отбора признаков в Weka предполагает совместное использование оценщика атрибутов и метода

поиска. Оценщик атрибутов – это метод, с помощью которого каждый признак оценивается в контексте целевой переменной. Метод поиска – это метод, с помощью которого просматривается признаковое пространство для поиска подходящего подмножества признаков.

В рамках данного исследования используются оценщики подмножества атрибутов CfsSubsetEval и WrapperSubsetEval. Первый оценщик атрибутов оценивает значимость подмножества признаков, учитывая индивидуальную прогностическую способность каждого признака, а также степень избыточности между ними. Второй – оцениваем эффективность подмножества признаков, учитывая результат примененного алгоритма классификации для оценки набора признаков. В поисковой стратегии может использоваться алгоритм, отличный от того, что будет затем использован для обучения классифицирующей модели.

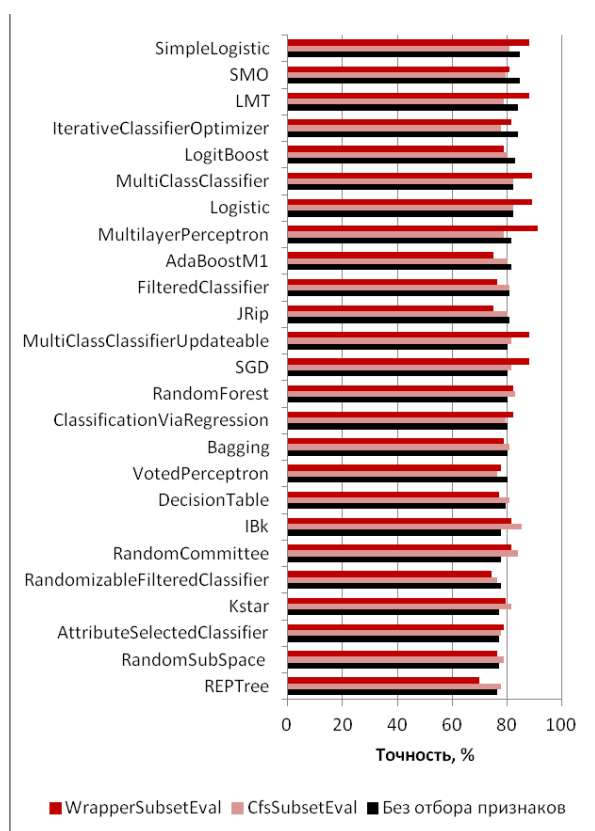


Рис. 1. Точность классификаторов

3. РЕЗУЛЬТАТЫ

В рамках эксперимента было отобрано 25 алгоритмов машинного обучения, входящих в пакет программ Weka, для которых точность классификации превысила 75 % при обучении на исходном признаковом пространстве (133 признака). Далее был произведен отбор признаков при помощи двух методов. Первый использует комбинацию оценщика атрибутов CfsSubsetEval и метода поиска GreedyStepwise, второй – комбинацию оценщика атрибутов WrapperSubsetEval с классификатором Logistic и метода поиска GreedyStepwise. На полученных подмножествах признаков вновь выполнено обучение классификаторов. Результаты показаны на Рис. 1.

При использовании оценщика атрибутов CfsSubsetEval было получено следующее подмножество признаков: активность Хьорта для канала O1, сложность Хьорта для канала F8, мощность дельта диапазона для канала Fp1, мощность тета диапазона для канала Fp1, мощность дельта диапазона для канала F8, мощность бета диапазона для канала C3 и мощность альфа диапазона для канала P4.

При использовании оценщика атрибутов WrapperSubsetEval были отобраны: активность Хьорта для канала T4, активность Хьорта для канала C4, сложность Хьорта для канала F8, мобильность Хьорта для канала F8, мобильность Хьорта для канала P4, мощность дельта диапазона для канала Fp1, мощность дельта диапазона для канала Fp2, мощность дельта диапазона для канала F8, мощность альфа диапазона для канала T6.

Отбор признаков первым методом оказался успешным для 12 алгоритмов классификации, вторым – для 13 алгоритмов. При этом наилучший результат первым методом был получен для алгоритма IBk (точность классификатора изменилась с 77,94 % до 85,29%). Наилучший результат вторым методом был получен для алгоритма MultilayerPerceptron (точность классификатора изменилась с 81,61% до 91,17 %).

4. ЗАКЛЮЧЕНИЕ

В данной работе исследованы входящие в пакет программ Weka методы отбора признаков, использующие оценщики подмножества атрибутов для сокращения размерности признакового пространства. В качестве примера для анализа выбрана задача классификации когнитивных состояний и состояний покоя по данным ЭЭГ. Дальнейшие исследования целесообразно связать с поиском оптимальных настроек для оценщиков подмножества атрибутов, учитывающих алгоритма классификации при оценке значимости признаков.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Министерства Науки и Высшего Образования Российской Федерации в рамках Госзадания (проект № FSUN-2020-0009).

ЛИТЕРАТУРА

- [1] Ahmad, R.F. Classification of cognitive and resting states of the brain using EEG features / R.F. Ahmad, A.S. Malik, H.U. Amin, N. Kamel, F. Reza // IEEE International Symposium on Medical Measurements and Applications (MeMeA). – 2016. – P. 1-5.
- [2] Mazher, M. Classification of resting and cognitive states using EEG-based feature extraction and connectivity approach / M. Mazher, I. Faye, A. Qayyum, A. S. Malik // IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). – 2018. – P.184-188.
- [3] Choi, S. Classification of mental arithmetic and resting-state based on Ear-EEG / S. Choi, G. Choi, H. Lee, H. Hwang, J. Shin // 6th International Conference on Brain-Computer Interface (BCI). – 2018. – P. 1-4.
- [4] Zyma, I. Electroencephalograms during mental arithmetic task performance / I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, O. Shpenkov // Data. – 2019. – Vol. 4(1). – P.14.
- [5] Cabañero-Gomez, L. Eeglib: A Python module for EEG feature extraction / L. Cabañero-Gomez, R. Hervás, I. González, L. Rodríguez-Benitez // SoftwareX. – 2021. – Vol. 15. – P. 100745.

Статистический анализ временных рядов температуры воздуха приземного слоя атмосферы в различных регионах России

В.А. Газарян
МГУ имени М.В. Ломоносова
Финансовый университет при
Правительстве РФ
Москва, Россия
vagazaryan@fa.ru

П.И. Кудрявцев
МГУ имени М.В. Ломоносова
Москва, Россия
pavel2195@gmail.com

А.В. Безрукова
МГУ имени М.В. Ломоносова
Москва, Россия
aleksandra_bezrukova@mail.ru

Ю.А. Курбатова
ИПЭЭ им. А.Н. Северцова РАН
Москва, Россия
kurbatova.j@gmail.com

Н.Е. Шапкина
МГУ имени М.В. Ломоносова
Москва, Россия
neshapkina@mail.ru

А.И. Чуличков
МГУ имени М.В. Ломоносова
Москва, Россия
achulichkov@gmail.com

Аннотация—Анализ региональных трендов приземной температуры воздуха приобрёл в последние годы особую актуальность в связи с изменениями климата, вызванными ростом средней глобальной температуры воздуха. В данной работе проводится оценка основных тенденций многолетней динамики и циклических изменений среднемесячной температуры воздуха в нескольких регионах России, анализируется статистическая значимость и адекватность построенных моделей, приводится их сравнительная характеристика, приводится прогноз региональных тенденций.

Ключевые слова— статистический анализ, временной ряд температуры воздуха, регрессионная модель, сезонная декомпозиция, Фурье-анализ, циклические колебания, проверка статистических гипотез.

1. ВВЕДЕНИЕ

В настоящее время изучение пространственно-временной изменчивости климатических параметров, в частности, температуры воздуха, является одной из актуальных задач, так как динамика данного показателя может как положительно, так и в отрицательно влиять на окружающую среду и человеческую деятельность. Статистический анализ и сравнительная характеристика основных тенденций и циклических изменений временных рядов среднесуточной температуры воздуха в приземном слое атмосферы проводятся на основе данных региональных метеорологических станций различных регионов России: Тверской области, Петрозаводска и Якутска (<http://aisori-m.meteo.ru/>).

2. АНАЛИЗ ОСНОВНЫХ ТЕНДЕНЦИЙ И ЦИКЛИЧЕСКИХ ОСОБЕННОСТЕЙ РЕГИОНАЛЬНЫХ ВРЕМЕННЫХ РЯДОВ ТЕМПЕРАТУРЫ ВОЗДУХА

Временной ряд температуры воздуха —упорядоченная совокупность его значений, измеряемых через строго постоянный промежуток времени [1]. Для построения модели временного ряда температуры в работе применяются такие методы как регрессионный, корреляционный и спектральный анализ [2], различные методы сглаживания и фильтрации [3], а также приводится интерпретация результатов моделирования и обсуждается прогноз региональных тенденций.

Поскольку за долгосрочный период наблюдений не выявлено заметного роста амплитуды сезонных колебаний, в работе используется аддитивная модель декомпозиции временного ряда

$$X_t = s_t + C_t + u_t + \varepsilon_t, \quad (1)$$

где s_t — сезонная компонента, C_t — циклическая компонента, u_t — тренд, определяющий основную тенденцию временного ряда, ε_t — случайная составляющая, $t = t_1, t_2, \dots, t_N$, N - число уровней ряда.

В результате декомпозиции региональных временных рядов температуры (1) построены полиномиальные регрессионные модели до 4-го порядка включительно, получены статистически значимые оценки параметров регрессии и эмпирические коэффициенты корреляции Пирсона при различных значениях интервала сглаживания.

Регрессионный анализ среднесуточной температуры воздуха за период 1963-2017 гг. в трех различных регионах России показал, что теснота линейной зависимости сглаженного временного ряда от времени связана с периодом сглаживания и достигает максимального значения коэффициента корреляции Пирсона 0,95 при восьмилетнем периоде сглаживания временных рядов температуры Тверской области и Петрозаводска, что позволяет выдвинуть предположение о наличии в указанных регионах восьмилетних температурных циклов.

Для оценки статистической значимости построенных полиномиальных регрессионных моделей решались задачи проверки статистических гипотез о равенстве нулю параметров уравнений регрессии на уровне значимости, не превышающим 0,01, а также задачи интервального оценивания параметров регрессии на уровне доверия не менее 0,9. Показано, что изменение периода сглаживания в диапазоне от 3 до 25 лет не оказывает существенного влияния как на результаты проверки статистических гипотез, так и на размеры доверительных интервалов, а вывод о значимости построенных регрессионных моделей временных рядов температуры подтверждается для всех регионов.

Значения коэффициента корреляции Пирсона для временных рядов Тверской области, Петрозаводска и Якутска составляют 0,60, 0,53 и 0,77 соответственно (рис. 1).

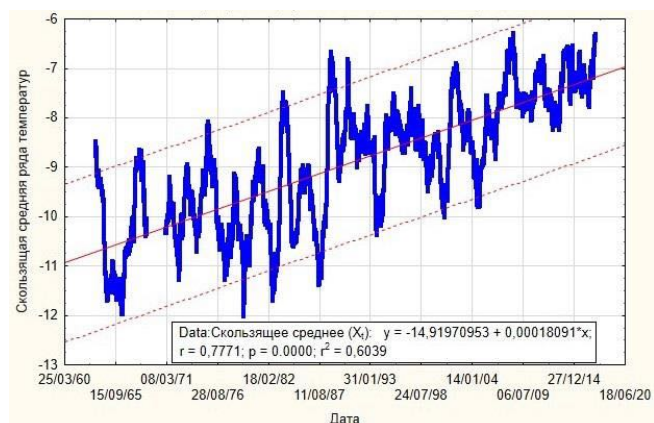


Рис. 1. Сглаженный временной ряд температуры воздуха (Якутск), график уравнения линейной регрессии (прямая линия), доверительный интервал уравнения регрессии на уровне доверия 0,9 (пунктирные линии)

Оценка адекватности построенных моделей результатам наблюдений проводилась с помощью проверки гипотезы о нормальном распределении остаточных компонент региональных рядов, а также о наличии в них линейной корреляционной связи. Критерии Колмогорова-Смирнова, Пирсона и авторегрессионный анализ показали, что остаточные составляющие во всех регионах имеют распределение, близкое к нормальному, что позволяет сделать вывод об адекватности построенных регрессионных моделей. Сравнение результатов проверки гипотез для трёх регионов показало, что наиболее качественной является регрессионная модель временного ряда температуры в Якутске.

После подтверждения стационарности региональных временных рядов с помощью теста Дики – Фуллера, был проведен Фурье-анализ сглаженных с годовым периодом временных рядов температуры, не содержащих сезонных компонент, оценен различный вклад выделенных циклических компонент в суммарные периодограммы сглаженных временных рядов в зависимости от региона. В результате подтверждено предположение о наличии восьмилетнего цикла во временных рядах Тверской области и Петрозаводска, выдвинутое при анализе влияния различных периодов сглаживания на тесноту линейной связи. Показано, что существенный вклад во временной ряд Петрозаводска вносит также циклическая компонента с периодом около трех лет, в то время как во временном ряду Якутска выявленные циклические компоненты вносят существенно меньший вклад в суммарную периодограмму.

Заключительная часть работы посвящена применению метода линейного прогноза, основанного на построении интегрированной модели авторегрессии ARIMA временных рядов температуры трех регионов [4]. Прогнозирование региональных временных рядов осуществлялось путем анализа автокорреляционной и частичной автокорреляционной функций, для оптимизации параметров моделей применялся коэффициент Акаике. Диагностическая проверка

подтвердила несмещенность, стационарность и неавтокоррелированность остаточных составляющих, что свидетельствует о значимости и адекватности построенных моделей прогноза динамики среднесуточной температуры воздуха.

3. ЗАКЛЮЧЕНИЕ

Проведенный регрессионный, корреляционный и спектральный анализ временных рядов приземной температуры воздуха трех регионов России показал, что тенденция увеличения среднесуточной температуры воздуха за полувековой период наблюдений отмечается во всех трех регионах, причём построенные регрессионные модели являются статистически значимыми и адекватными. Таким образом, рост среднегодовой температуры воздуха за 1963–2017 гг. составил 2,4°C, 2,1°C и 3,4°C для Тверской области, Петрозаводска и Якутска соответственно. Полученные температурные тренды согласуются с наблюдаемым ростом средней глобальной температуры воздуха. Наиболее значимой и адекватной из построенных в данной работе регрессионных моделей для сглаженных с годовым периодом временных рядов является линейный тренд среднесуточной температуры воздуха в Якутске с коэффициентом корреляции Пирсона, равным 0,77 (рис. 1). Что касается спектрального анализа, то основной вклад во временные ряды Тверской области и Петрозаводска вносят гармоники около восьми и трех лет, которые определяются, по-видимому, особенностями атмосферной циркуляции и солнечно-земных связей, в частности, 11-летним циклом Швабе-Вольфа, в котором за первые 3-4 года происходит увеличение числа солнечных пятен и усиление других проявлений солнечной активности, а в течение последующих 7-8 лет – обратный процесс. При анализе циклических колебаний следует отметить особенность ряда температуры в Якутии: в отличие от временных рядов Тверской области и Петрозаводска, существенных циклических составляющих при построении аддитивной модели временного ряда Якутска не выявлено. По-видимому, такие результаты связаны с особенностями атмосферной циркуляции на региональном уровне.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке РФФИ, проект № 19-29-09044.

ЛИТЕРАТУРА

- [1] Gazaryan, V.A. Contemporary climate changes in the southwest of the valdai hills: A statistical analysis of the long-term dynamics of the air temperature / V.A. Gazaryan, J.A. Kurbatova, T.A. Ovsyannikov, N.E. Shapkina // Moscow University Physics Bulletin. – 2015. – Vol. 70(5). – P. 346-352.
- [2] Gazaryan, V.A. A statistical analysis of cyclical changes in the time series of meteorological parameters in the southwest of the valdai hills / V.A. Gazaryan, Y.A. Kurbatova, T.A. Ovsyannikov, N.E. Shapkina // Moscow University Physics Bulletin. – 2018. – Vol. 73(1). – P. 61-67.
- [3] Avilov, V.K. Morphological and other research techniques for almost cycloime series as applied to CO2 concentration series / V.K. Avilov, V.S. Aleshnovskii, A.V. Bezrukova // Computational Mathematics and Mathematical Physics. – 2021. – T. 61, № 7. – С. 1106-1117.
- [4] Jain, E. A Study of time series models ARIMA and ETS / E. Jain, D. Mallick // International Journal of Modern Education and Computer Science (IJMECS). – 2017. – Vol. 9(4). – P. 57-63.

Алгоритм оценки продолжительности стадий заболевания по набору неполных данных

П.Н. Подзолков
Тюменский государственный университет
Тюмень, Россия
p.n.podzolkov@utmn.ru

Аннотация—В статье рассматриваются вопросы математического моделирования развития заболеваний. Поднимается проблема оценки временных показателей по неполным данным для построения моделей многостадийных процессов. Предложен алгоритм выравнивания набора последовательностей, содержащих неполную информацию о течении заболевания, для оценки продолжительностей его стадий. Представлены результаты тестирования алгоритма на симулированных данных.

Ключевые слова— моделирование заболеваний, выравнивание, неполные данные, компартментальная модель эпидемии, цепь Маркова.

1. ВВЕДЕНИЕ

Хронические и инфекционные заболевания до сих пор остаются значительной проблемой мирового здравоохранения. Существуют различные методы математического моделирования и прогнозирования развития заболеваний. Большинство методов основано на представлении заболевания в виде многостадийного процесса, проходя через который, пациент переходит последовательно из одной стадии в другую. В соответствии с работой Tolles J. и Luong T. В. по такому механизму работают компартментальные модели эпидемий. Андреев Д. М. показал, что модели хронических заболеваний, применяющие цепи Маркова, основаны на том же принципе.

Одним из основных этапов построения подобных моделей является вычисление средней продолжительности каждой стадии и соответственно оценка вероятности перехода в следующую стадию в каждый момент времени. Такие показатели могут быть вычислены по результатам наблюдений за пациентами, проходившими через процесс данного заболевания. Однако в большинстве случаев есть информация только о нескольких, проведённых с различными интервалами обследованиях пациента. Согласно Михальскому А. И. подобная ситуация соответствует интервально-цензурированным данным. Результатом каждого обследования является определённая стадия заболевания, однако в промежутках между обследованиями данных о состоянии индивида нет и, следовательно, момент перехода в следующее состояние, если таковой произошёл, не известен.

Не имея установленных моментов переходов между стадиями, нельзя определить длительность протекания каждой стадии заболевания и соответственно оценить вероятность перехода из неё в конкретный момент времени. Для решения этой проблемы предлагается алгоритм выравнивания последовательностей, описывающих прохождение заболевания у индивидов.

2. КРИТЕРИИ АЛГОРИТМА

На вход алгоритму подаются строки, содержащие информацию о пройденных состояниях, разделённых интервалами различной длины, которые соответствуют промежуткам времени между обследованиями. Задачей алгоритма является определение смещения для каждой последовательности, такое, чтобы моменты наступления и завершения стадий максимально соответствовали другим последовательностям.

Для определения критериев алгоритма были введены следующие понятия. Гарантированный промежуток стадии (ГПС) – промежуток времени в последовательности наблюдений, который начинается с первым вхождением данной стадии и заканчивается последним вхождением данной стадии. Допустимый промежуток стадии (ДПС) – промежуток времени в последовательности наблюдений, который начинается после завершения ГПС предыдущей наблюдаемой стадии и заканчивается перед ГПС следующей наблюдаемой стадии. Примеры определения указанных промежутков выделены на Рис. 1.

Критерием качества выравнивания в таком случае становится максимальное соответствие ДПС и ГПС каждой стадии у последовательности с данными промежутками соответствующей стадии другой последовательности.

ГПС стадии b	a	-	b	-	-	b	-	c
ДПС стадии b	a	-	b	-	-	b	-	c

Рис. 1. Пример последовательности с выделенными серым цветом промежутками стадии b

3. ПАРНОЕ ВЫРАВНИВАНИЕ

Для определения качества конкретного взаимного расположения двух последовательностей вводится штраф за расстояние между серединами ДПС и ГПС соответствующих стадий двух последовательностей.

Во многих случаях для конкретной стадии нельзя определить границы ДПС или ГПС. В случае, если нет информации о предыдущих или последующих состояниях пациента, нельзя определить ДПС выбранной стадии. А если стадия ни разу не наблюдалась у пациента, то нельзя определить её ГПС. Таким образом, необходим набор правил, по которым будет находится штраф между двумя выравниваемыми последовательностями. В таблице 1 указан вариант правил для оценки штрафа расстояний. В случае, если у одной из последовательностей невозможно определить ни один промежуток для выбранной стадии, то эта стадия не влияет на итоговый штраф выравнивания выбранной пары последовательностей. Для вычисления штрафа каждой стадии необходимо найти сумму квадратов расстояний, указанных в Таблице 1. Штраф каждой

стадии прибавляется к общему счётику штрафа выравнивания двух последовательностей. Таким образом, для каждого варианта смещения двух последовательностей вычисляется величина характеризующая неудовлетворительность данного варианта. Варианты с наименьшим штрафом выбираются как наиболее корректные.

Таблица 1. Правила вычисления расстояний между центрами промежутков конкретной стадии в двух последовательностях

Интервалы, определённые во 2ой последовательности	Интервалы, определённые в 1ой последовательности		
	ГПС	ДПС	ГПС и ДПС
ГПС	$\text{dist}^a(g_1^b, g_2^c)$	$\text{dist}(d_1^d, g_2)$	$\text{dist}(g_1, g_2)$
ДПС	$\text{dist}(g_1, d_2^e)$	$\text{dist}(d_1, d_2)^c$	$\text{dist}(g_1, d_2)$ $\text{dist}(d_1, d_2)$
ГПС и ДПС	$\text{dist}(g_1, g_2)$	$\text{dist}(d_1, g_2)$; $\text{dist}(d_1, d_2)$	$\text{dist}(g_1, g_2)$; $\text{dist}(d_1, d_2)$

^a Расстояние между указанными точками; b. Центр ГПС в 1 последовательности; c. Центр ГПС во 2 последовательности; d. Центр ДПС в 1 последовательности; e. Центр ДПС во 2 последовательности

4. МНОЖЕСТВЕННОЕ ВЫРАВНИВАНИЕ

Выравнивание набора последовательностей происходит посредством последовательного попарного выравнивания различных последовательностей из набора. Перед этим набор сортируется в порядке уменьшения количества присутствующих в последовательностях обследований. Затем выполняется парное выравнивание двух первых последовательностей. Из всех вариантов их выравнивания выбираются первые N выравниваний в списке, отсортированном по увеличению штрафа. N – «память алгоритма», величина определяемая пользователем, увеличение которой улучшает итоговое множественное выравнивание, но ухудшает производительность алгоритма.

Затем к выравниванию добавляется следующая последовательность. Для этого для каждого из N выбранных выравниваний перебираются все возможные смещения новой последовательности и вычисляются соответствующие штрафы. Штраф для новой последовательности вычисляется как сумма штрафов её текущего положения относительно некоторых последовательностей, добавленных к выравниванию ранее. Среди всех положений новой последовательности во всех N существующих выравниваниях снова выбирается N лучших, то есть имеющих минимальный штраф. Так происходит до тех пор, пока не будут добавлены все последовательности. В результате работы алгоритма множественного выравнивания получаем N лучших выравниваний. В каждом выравнивании описаны целочисленные смещения каждой последовательности.

5. ТЕСТИРОВАНИЕ

Тестирование алгоритма проводилось на симулированных данных. Симуляция данных заключалась в определении параметров (математическое ожидание и стандартное отклонение) для каждой стадии. По этим параметрам формировался набор последовательностей, в которых длительность каждой стадии определялась как величина из нормального распределения с выбранными для этой стадии параметрами. Далее производилось сокрытие информации в каждой последовательности с сохранением только нескольких позиций, имитирующих моменты

обследований и расположенных с определённым интервалом, длительность которого также соответствовала нормально распределённой величине с заданными параметрами.

На Рис. 2 приведён пример выравнивания 10 последовательностей процесса из 3 стадий (a, b, c). Математические ожидания продолжительности стадий при симуляции были выбраны следующие (стандартное отклонение у всех стадий задано 1): a – 10, b – 3, c – 6. Моменты обследований выбирались со средним интервалом в 7 единиц при стандартном отклонении равном 3.

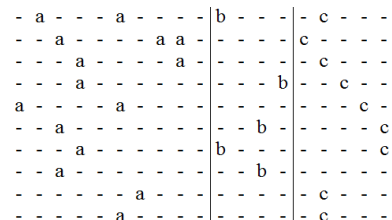


Рис. 2. Пример выравнивания 10 последовательностей

По результатам выравнивания можно оценить продолжительности стадий: a – от 9 до 10, b – от 4 до 5, c – около 5. Такие оценки в значительной мере соответствуют задаваемым исходным значениям при симуляции.

6. ЗАКЛЮЧЕНИЕ

Нами разработан подход к оценке продолжительности стадий заболевания по набору интервально-цензурированных данных. Планируется дальнейшая доработка описанного выше алгоритма, но уже сейчас можно заметить потенциал данного подхода к оценке длительности интервалов в неполных данных. Представленный подход применим к оценке временных показателей любых многостадийных процессов.

ЛИТЕРАТУРА

- [1] Martino, A. Multivariate hidden markov models for disease progression / A. Martino, G. Guatteri, A.M. Paganoni // Statistical Analysis and Data Mining: The ASA Data Science Journal. – 2020. – Vol. 13(5). – P. 499-507.
- [2] Tolles, J. Modeling epidemics with compartmental models / J. Tolles, T.B. Luong // Jama. – 2020. – Vol. 323(24). – P. 2515-2516.
- [3] Андреев, Д.М. Стандартизация моделирования прогрессирования хронических заболеваний / Д.А. Андреев, Н.В. Хачанова, В.Н. Степанова, Е.Е. Башлакова, Е.П. Евдошенко, М.В. Давыдовская // Проблемы стандартизации в здравоохранении. – 2017. – № 9-10. – С. 12-23.
- [4] Кондратьев, М.А. Методы прогнозирования и модели распространения заболеваний / М.А. Кондратьев // Компьютерные исследования и моделирование. – 2013. – Т. 5, № 5. – С. 863-882.
- [5] Маркович, Н.М. Оценка эпидемиологических показателей заболеваемости по косвенным данным / Н.М. Маркович, А.И. Михальский, В. Моргенштерн // Автоматика и телемеханика. – 1998. – № 6. – С. 153-162.
- [6] Михальский, А.И. Моделирование заболеваний как обратная задача / А.И. Михальский // Труды 3-й Международной конференции «Высокие технологии, исследования, образование в физиологии, медицине и фармакологии». – СПб.: Издательство Политехнического университета, 2012. – Т. 2. – С. 213-216.
- [7] Романюха, А.А. Математические модели в иммунологии и эпидемиологии инфекционных заболеваний / А.А. Романюха. – Москва: Бином. Лаборатория знаний, 2018. – 293 с.

Программный инструментарий нейрокомпьютерного интерфейса для анализа электроэнцефалограмм на базе непрерывного вейвлет-преобразования

А.А. Столбова
Самарский национальный
исследовательский университет
им. академика С.П. Королева
Самара, Россия
stolbova.aa@ssau.ru

В.Ю. Дьяконов
Самарский национальный
исследовательский университет
им. академика С.П. Королева
Самара, Россия
tradeincorp2406@gmail.com

О.К. Головнин
Самарский национальный
исследовательский университет
им. академика С.П. Королева
Самара, Россия
golovnin@ssau.ru

Аннотация—В статье представлен разработанный программный инструментарий нейрокомпьютерного интерфейса для анализа электроэнцефалограмм на базе непрерывного вейвлет-преобразования. Применение программного инструментария позволяет осуществлять вейвлет-анализ сигналов и фиксировать изменения частоты сигнала во времени, что дает возможность осуществлять поиск и выявление деталей и аномалий, развивающихся во времени.

Ключевые слова—нейрокомпьютерный интерфейс, вейвлет-анализ, когерентность, электроэнцефалограмма, ЭЭГ.

1. ВВЕДЕНИЕ

В настоящее время активно ведутся исследования в области разработки нейрокомпьютерных интерфейсов, с помощью которых реализуется управление протезами, экзоскелетами и роботами, применяющимися в реабилитационной медицине и профессиональном спорте. Люди с тяжелыми неврологическими нарушениями сталкиваются со многими проблемами в сенсомоторных функциях и взаимодействии с окружающей средой. Так, серьезной проблемой остается нарушение походки и равновесия после инсультов и травм головного мозга а также последующее восстановление повседневной и социальной активности.

При разработке нейрокомпьютерных интерфейсов, используемых при решении задач реабилитации пациентов, в качестве неинвазивных портативных методов нейровизуализации применяется электроэнцефалография (ЭЭГ). Так, например, применительно к задаче реабилитации, в [1] предложена система, построенная на основе 3D-игры, управляемой нейрокомпьютерным интерфейсом, а в [2] авторы проводят исследование ускоренного восстановления пациентов после инсульта, достигаемого за счет использования интерфейса «мозг-компьютер» на основе анализа данных ЭЭГ.

Нейрокомпьютерные интерфейсы, основанные на образах движения, позволяют людям использовать мысленные моторные образы, распознаваемые по электроэнцефалограмме, в качестве управляющих команд, даже без использования элементов визуального контроля. Поскольку нестационарный характер сигналов ЭЭГ делает нейрокомпьютерный интерфейс на основе

ЭЭГ динамической системой [3], в качестве основного метода анализа предлагается использовать вейвлет-преобразование, позволяющее локализовать спектральные характеристики сигнала во времени. Применение вейвлет-преобразования позволяет получить преимущества при решении большого класса биомедицинских задач и, в частности, при анализе нестационарных сигналов, к которым и относятся сигналы ЭЭГ [4].

В статье предлагается программный инструментарий нейрокомпьютерного интерфейса для анализа электроэнцефалограмм на базе непрерывного вейвлет-преобразования.

2. АЛГОРИТМ ВЕЙВЛЕТ-АНАЛИЗА СИГНАЛОВ ЭЭГ

Разработан алгоритм вейвлет-преобразования, предназначенный для использования в программном инструментарии нейрокомпьютерного интерфейса. Алгоритм продолжает изыскания, выполненные в [5], и позволяет выполнить обработку сигналов ЭЭГ, поступающих с многоканальных устройств. Алгоритм представлен следующим набором шагов, выполняемых для всех каналов ЭЭГ.

Шаг 1. Получить массив масштабов:

$$a_i = \frac{1}{w_{\min} + i \times \Delta w}, \quad i = 0, \dots, N_a - 1, \quad (1)$$

где a_i – значение i -го масштаба, w_{\min} – минимальная частота, Δw – интервал дискретизации частоты, N_a – число масштабов.

Шаг 2. Рассчитать значения масштабированных вейвлетов при равном нулю сдвиге для каждого значения масштаба. Для этого выполняются следующие шаги.

Шаг 2.1. Определить количество отсчетов вейвлета для текущего масштаба a_i , попадающих в ширину w_i :

$$n_i = \text{ent} \left[\frac{w_i}{\Delta t_0} + 0,5 \right], \quad (2)$$

где Δt_0 – интервал дискретизации вейвлета, $\text{ent}[\]$ – операция взятия целой части.

Шаг 2.2. Вычислить значение вейвлета:

$$\Psi_{i,k} = \frac{1}{\sqrt{a_i}} \Psi\left(\frac{k\Delta t_0}{a_i}\right), \quad (3)$$

где $k = -ent\left[\frac{n_i}{2}\right] \dots ent\left[\frac{n_i}{2}\right]$.

Шаг 2.3. Повторить шаги 2.1 и 2.2 для каждого i -ого масштаба.

Шаг 3. Вычислить коэффициенты вейвлет-преобразования, например, следующим способом:

$$W(a,b) = \frac{1}{\sqrt{a}} \sum_{i=0}^{N-1} x(t_i) \psi\left(\frac{t_i-b}{a}\right), \quad (4)$$

где ψ – материнский вейвлет, a – масштаб, b – временной сдвиг, t_i – i -ый временной отсчет, $x(t_i)$ – i -ый отсчет сигнала, N – число отсчетов в сигнале.

3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ И РЕЗУЛЬТАТЫ

Программный инструментальный нейрокомпьютерного интерфейса разработан на языке C# с использованием технологии Windows Presentation Foundation для визуализации данных. Проведено исследование с использованием разработанного программного инструментария для выявления зависимости между физическими и мысленным моторными движениями человека. В процессе регистрации ЭЭГ испытуемые выполняли физическое и мысленное моторное движение кисти руки или одного пальца руки. Во время эксперимента давалась голосовая команда, после которой выполнялось одно из предложенных действий. При регистрации ЭЭГ использовались следующие электроды: Fp1A1, Fp2A2, T3A1, T4A2, C3A1, C4A2, Cz. Таким образом, исходный сигнал ЭЭГ представляет собой набор данных об активности мозга для 7 электродов, регистрируемых с частотой дискретизации 500 Гц (Рис. 1).

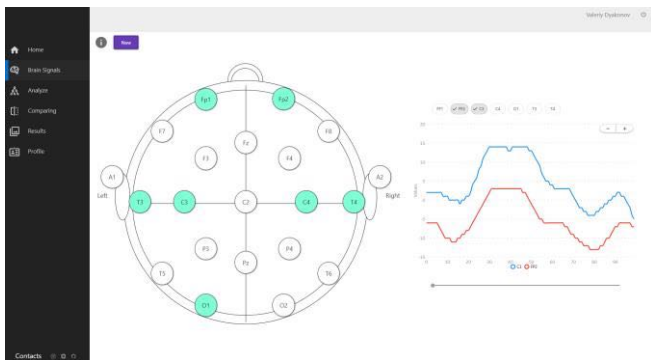


Рис. 1. Программный инструментальный нейрокомпьютерного интерфейса

В качестве анализирующего вейвлета выбрана восьмая производная Гауссовой функции с максимальным масштабом 500. Отметим, что масштаб обратен частоте, а, следовательно, меньший масштаб соответствует более высоким частотам, и наоборот.

В процессе проведения исследования вычислялась когерентность вейвлетов для обнаружения общих

локализованных временем колебаний (Рис. 2). Для численного выражения меры сравнения вычислены среднее арифметическое и среднеквадратическое отклонение.

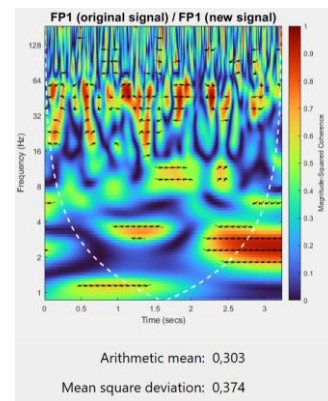


Рис. 2. Результат вычисления когерентности для сигналов FP1

Анализ полученных результатов вычисления когерентности для всех сигналов ЭЭГ позволяет сделать следующие выводы: осуществляется фиксирование изменений частоты сигнала во времени; сигналы, регистрируемые в различных каналах, схожи по своему поведению; при использовании воображения испытуемым сигналы более активны на больших масштабах; наибольшая схожесть сигналов приходится на диапазон частот 32–64 Гц; сигналы схожи приблизительно на 29–35%.

4. ЗАКЛЮЧЕНИЕ

Таким образом, разработан программный инструментальный нейрокомпьютерного интерфейса, предназначенный для анализа электроэнцефалограмм с использованием непрерывного вейвлет-преобразования. Вейвлет-анализ сигналов, поступающих с различных каналов устройств ЭЭГ, позволяет фиксировать изменения частоты сигнала во времени, что, в свою очередь, позволяет осуществлять поиск и выявление деталей и аномалий, развивающихся во времени.

ЛИТЕРАТУРА

- [1] Aamer, A. BCI Integrated with VR for Rehabilitation / A. Aamer, A. Esawy, O. Swelam, T. Nabil, A. Anwar, A. Eldeib // 31st International Conference on Microelectronics. – 2019. – P. 166-169.
- [2] Foong, R. Assessment of the efficacy of EEG-based MI-BCI with visual feedback and EEG correlates of mental fatigue for upper-limb stroke rehabilitation / R. Foong, K.K. Ang, C. Quek, C. Guan, K.S. Phua, C.W.K. Kuah, K.S.G. Chua // IEEE Transactions on Biomedical Engineering. – 2019. – Vol. 67(3). – P. 786-795.
- [3] Raza, H. Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface / H. Raza, D. Rathee, S.M. Zhou, H. Cecotti, G. Prasad // Neurocomputing. – 2019. – Vol. 343. – P. 154-166.
- [4] Khan, N.A. A new feature for the classification of non-stationary signals based on the direction of signal energy in the time–frequency domain / N.A. Khan, S. Ali // Computers in biology and medicine. – 2018. – Vol. 100. – P. 10-16.
- [5] Stolbova, A. Wavelet-Based Arrhythmia Detection in Medical Diagnostics Sensor Networks / A. Stolbova, S. Prokhorov, A. Kuzmin, A. Ivaschenko // International Conference on Information Technologies. – 2019. – P. 468-479.

Прогнозирование состояния технического объекта на основе модели системы квазипериодических процессов в виде изображений на цилиндре

В.Р. Крашенинников
Ульяновский государственный
технический университет
Ульяновск, Россия
kvrulstu@mail.ru

Ю.Е. Кувайскова
Ульяновский государственный
технический университет
Ульяновск, Россия
u.kuvaiskova@mail.ru

В.Н. Клячкин
Ульяновский государственный
технический университет
Ульяновск, Россия
v_kl@mail.ru

Аннотация—Часто техническое состояние объекта характеризуется несколькими диагностируемыми параметрами его функционирования, то есть системой взаимосвязанных временных рядов, в динамике которых наблюдается периодичность со случайными изменениями периодов – квазипериодичность, не поддающаяся точному прогнозу. При этом для оценивания и прогнозирования состояния объекта требуется построение модели такой системы процессов и её идентификация для конкретного объекта по результатам его наблюдений. В данной работе для решения этих задач предлагается использовать модель системы квазипериодических процессов в виде развёрток нескольких цилиндрических изображений по спирали. Эффективность модели демонстрируется на примере прогнозирования состояния системы водоочистки.

Ключевые слова— прогнозирование, технический объект, система временных рядов, квазипериодический процесс, модель цилиндрического изображения

1. ВВЕДЕНИЕ

Задача прогнозирования технического состояния объекта на предстоящий временной период играет очень важную роль в обеспечении его стабильной работы [1]. Состояние многих технических объектов описывается набором диагностируемых параметров, представляющих собой систему временных рядов, в динамике которых наблюдается нерегулярная периодичность – квазипериодичность, то есть повторения с компонентом непредсказуемости, не поддающимся точному прогнозу. Например, вибрация различных двигателей, турбин, гидроагрегатов. В настоящей работе для прогнозирования состояния технических объектов предлагается использовать модели систем квазипериодических процессов в виде изображений, определённых на «многослойных» или «толстостенных» цилиндрах. Каждый слой цилиндра соответствует одному из временных рядов системы, корреляции между которыми задаются «внешним» уравнением авторегрессии, что позволяет описывать различные зависимости между процессами системы.

2. МОДЕЛЬ СИСТЕМЫ КВАЗИПЕРИОДИЧЕСКИХ ПРОЦЕССОВ В ВИДЕ ИЗОБРАЖЕНИЙ НА ЦИЛИНДРЕ

Квазипериодический процесс отличается наличием двойной коррелированности: имеется значимая корреляция как между соседними отсчётами, так и между отсчётами, отстоящими на несколько периодов. Для

описания таких процессов будем использовать авторегрессионные модели цилиндрических изображений [2]. Для построения таких моделей рассмотрим спиралевидную сетку на цилиндре, строки которой представляют собой витки спирали. Тогда цилиндрическое изображение можно рассматривать как одну последовательность отсчётов вдоль этой спирали.

Рассмотрим аналог авторегрессионной модели Хабиби [3]:

$$x_{k,l} = s x_{k,l-1} + r x_{k-1,l} - s r x_{k-1,l-1} + q \xi_{k,l}, \quad (1)$$

где k – номер витка спирали; l – номер узла в витке; s и r – параметры модели; $\xi_{k,l}$ – независимые стандартные случайные величины; при этом $l = 0, T-1$; $x_{k,l} = x_{k+1,l-T}$ при $l \geq T$; T – период, то есть количество точек в одном витке. Для использования модели (1) значения на первом витке должны быть заданы отдельно.

Если представить модель цилиндрического изображения в виде развёртки изображения вдоль спирали, то получим модель квазипериодического случайного процесса:

$$x_n = s x_{n-1} + r x_{n-T} - s r x_{n-T-1} + q \xi_n, \quad (2)$$

где $n = kT + l$.

Значение параметра s оказывает влияние на гладкость процесса, то есть на коррелированность отсчётов вдоль строк, а параметр r – на коррелированность отсчётов на расстоянии периода. При этом значения параметра r близкие к единице показывают, что соседние строки изображения (витки спирали) будут сильно коррелированы, поэтому данную модель можно использовать для описания поведения квазипериодических временных рядов.

Для описания динамики системы временных рядов рассмотрим модель, в которой каждый временной ряд системы представлен изображением на своём цилиндре, а эти отдельные цилиндры вложены друг в друга и образуют систему цилиндров – «многослойный» или «толстостенный» цилиндр.

Зададим систему из m однородных случайных процессов $X = \{x_{n,m}, m=1,2,\dots,M\}$, каждый из которых описывается с помощью модели (2).

Тогда ковариационная матрица между процессами будет иметь вид:

$$C = \{C_{ij}\} = \{\text{cov}(x_{n,i}, x_{n,j})\} = \{M[x_{n,i}x_{n,j}]\}, \quad (3)$$

$$i = 1, 2, \dots, M, j = 1, 2, \dots, M$$

Пусть $Y = \{y_{n,m} \ m=1, 2, \dots, M\}$ – система из m независимых «стандартных» процессов с нулевым средним и единичной дисперсией, полученных на основе модели (2), в которой параметр $q^2 = q_1^2 = 1/\sigma^2$, где σ^2 – дисперсия процесса:

$$y_{n,m} = s y_{n-1,m} + r y_{n-T,m} - s r y_{n-T-1,m} + q_1 \xi_{n,m}. \quad (4)$$

Тогда заданную систему $X = \{x_{n,m} \ m=1, 2, \dots, M\}$ образуем в виде линейной комбинации «стандартных» процессов: $X = AY$, то есть

$$x_{n,m} = \sum_j a_{mj} y_{n,j}, \quad m = 1, 2, \dots, M, \quad (5)$$

где матрица A должна удовлетворять уравнению $AA^T = C$ и может быть взята нижней треугольной [2].

В частности, если $C_{ij} = \sigma^2 p^{|j-i|}$, то есть процессы $x_{n,m}$ имеют одинаковую дисперсию σ^2 и коэффициент корреляции $p^{|j-i|}$ между процессами $x_{n,i}$ и $x_{n,j}$, то они могут быть получены из стандартных в виде авторегрессии первого порядка:

$$x_{n,1} = \sigma y_{n,1}, \quad x_{n,m} = p x_{n,m-1} + \sqrt{1-p^2} y_{n,m}. \quad (6)$$

Переходя в формулах (3-6) от обозначений $x_{n,m}$ из модели (2) к обозначениям $x_{k,l,m}$ из модели (1), получим систему изображений, определённых на системе вложенных друг в друга цилиндров.

Таким образом, задавая значения параметров модели можно описывать системы квазипериодических процессов с требуемыми корреляциями внутри и между периодами, а также между процессами системы.

3. ЧИСЛЕННЫЙ ЭКСПЕРИМЕНТ

В качестве объекта исследования рассмотрим систему водоочистки в г. Санкт-Петербурге, Россия [1]. Техническое состояние системы оценивалось по параметрам качества питьевой воды в зависимости от физико-химических показателей водоисточника: X_1 – температура, X_2 – цветность, X_3 – мутность, X_4 – pH, X_5 – щёлочность, X_6 – окисляемость, X_7 – аммоний, X_8 – хлориды. Всего получено 348 наблюдений по каждому параметру системы восьми временных рядов.

Для оценки эффективности предлагаемого подхода разделим исходную выборку данных на две части: обучающую и контрольную. По обучающей части построим три модели системы временных рядов: модель авторегрессии проинтегрированного скользящего среднего (АРПСС) [4], модель векторной авторегрессии [5] и модель системы квазипериодических процессов в виде изображений на цилиндре.

Контрольную часть выборки будем использовать для оценки точности прогнозирования моделей. Для этого рассчитаем среднеквадратическое отклонение:

$$\sigma_{\Delta} = \sqrt{\sum_{t=1}^k (x_t - \hat{x}_t)^2 / (k - p)}, \quad (7)$$

где x_t – результат t -го наблюдения, \hat{x}_t – прогнозируемое значение по построенной модели, k – объём контрольной выборки; p – число оцениваемых параметров в модели.

Значения среднеквадратических отклонений, полученных по контрольной выборке, для каждого ряда системы представим в таблице 1.

ТАБЛИЦА 1. Точность прогнозирования системы временных рядов

Ряд	Точность прогнозирования (σ_{Δ})		
	АРПСС	Векторная авторегрессия	Модель системы квазипериодических процессов в виде изображений на цилиндре
X_1	1,75	0,75	0,56
X_2	2,31	1,91	1,86
X_3	6,01	5,85	5,44
X_4	0,19	0,11	0,10
X_5	0,09	0,04	0,03
X_6	1,07	0,68	0,65
X_7	0,21	0,11	0,09
X_8	58,39	56,02	51,37

Полученные результаты показывают, что использование модели системы квазипериодических процессов в виде развёрток нескольких цилиндрических изображений по спирали при прогнозировании системы физико-химических показателей водоисточника позволяет повысить точность прогнозирования до 3 раз по сравнению с моделью АРПСС и до 1,3 раза по сравнению с моделью векторной авторегрессии.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-01-00613.

ЛИТЕРАТУРА

- [1] Клячкин, В.Н. Прогнозирование и диагностика стабильности функционирования технических объектов: монография / В.Н. Клячкин, В.Р. Крашенинников, Ю.Е. Кувайскова. – М.: РУСАЙНС, 2020. – 200 с.
- [2] Krashennnikov, V.R. Spiral Autoregressive Model of a Quasiperiodic Signal / V.R. Krashennnikov, D.V. Kalinov, Yu.G. Pankratov // Pattern Recognition and Image Analysis. – 2001. – Vol. 11(1). – P. 211-213.
- [3] Хабиби, А. Двумерная байесовская оценка изображений / А. Хабиби // ТИИЭР. – 1972. – Т. 60, № 7. – С. 153-159.
- [4] Бокс, Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – 242 с.
- [5] Sims, C.A. Macroeconomics and Reality / C.A. Sims // Econometrica. – 1980. – Vol. 48. – P. 1-48.

Алгоритмическое обеспечение персонального виртуального ассистента для автоматизации обработки клиентских запросов

К.С. Добратулин
НИТУ «МИСиС»
Москва, Россия
dobratulin@yahoo.com

М.И. Нежурина
НИТУ «МИСиС»
Москва, Россия
min@misis.ru

Аннотация—В данной статье описывается процесс создания алгоритмического обеспечения для функционирования персонального виртуального ассистента, позволяющего автоматизировать обработку запросов клиентов. Целью исследования является сокращение ошибок и времени обработки запроса клиента в бизнес-системах – текстовых чатах или голосовых каналах с использованием системы транскрибации текста. Приводятся результаты разработки алгоритмического обеспечения и оценка качества работы на синтетических данных.

Ключевые слова— виртуальный ассистент, обработка естественного языка, автоматизация обработки запросов, алгоритмическое обеспечение.

1. ВВЕДЕНИЕ

Персональные виртуальные ассистенты прошли путь развития длиной в несколько десятилетий – от систем распознавания и транскрибации текста до многофункциональных программных агентов с применением технологических достижений из области искусственного интеллекта [1]. В 2022 году персональные виртуальные ассистенты выполняют задачи для конечного пользователя и используются в бизнес-сервисах, таких как: виртуальные операторы, системы автоматического ответа, умные устройства, голосовые ассистенты, интерактивное голосовое меню [2]. Персональные виртуальные ассистенты, ориентированные на конечного потребителя, внедряются в программные комплексы бизнес-систем. Использование персональных виртуальных ассистентов в системах автоматизированной обработки клиентских запросов позволяет бизнесу строить новые сценарии взаимодействия с пользователем, улучшать показатели бизнес-метрик на основе оптимизации ресурсов и влияния на факторы технологические, временные, экономические и поведенческие [3]. Такие возможности использования послужили обоснованием необходимости алгоритмического обеспечения персонального виртуального ассистента.

2. ОПИСАНИЕ ЛИНГВИСТИЧЕСКИХ ПРИЗНАКОВ

Лингвистические признаки – основные при классификации запроса клиента. Так как специфика взаимодействия с персональным виртуальным ассистентом подразумевает отправку пользователем запроса в текстовом или голосовом канале, становится возможным на этапе обработки запроса выделить признаки из текстового представления запроса пользователя. Такими признаками являются признаки

лексические, текстовые и лексико-грамматические [4]. По таким признакам становится возможно классифицировать пользовательские данные, например, присвоив некоторую оценку сложности запроса на основе лингвистических признаков, попытаться проанализировать настроение пользователя на основе употребляемых в контексте запроса слов или фраз, удалить повторы или незначимые слова по текстовым признакам для статистического анализа. При разработке персонального виртуального ассистента следует внимательно учитывать потенциал анализа лингвистических признаков, так как это позволяет применять подходы из области искусственного интеллекта [5].

3. АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Для прогнозирования цели клиентского запроса на основе текста пользователя, в данной работе используются модели машинного обучения – логистическая регрессия и библиотека fastText, разработанная Facebook [6, 7]. Выбранные модели хорошо показали себя при выполнении задач, связанных с классификацией текста [8]. Также выбор в пользу данных моделей был сделан по причине их быстрой работы и сравнительно низких требований к вычислительным ресурсам, что является важными факторами при промышленной эксплуатации [9].

В качестве языка программирования для интеграции компонентов загрузки данных и алгоритмического обеспечения был выбран язык Python [10]. Строгих ограничений по техническому обеспечению для применения выбранных технологий в данной статье не заявлено, они зависят от количественной нагрузки пользовательских запросов и должны подбираться индивидуально, исходя из задач. В качестве результата сравнения моделей, помимо качества классификации пользовательских запросов алгоритмическим обеспечением, получены время выполнения классификации одного запроса и размер модели, занимаемый на жестком диске

4. РЕЗУЛЬТАТЫ

По результатам работы прототипа получены метрики на подготовленных наборах данных. Выборкой из синтетических данных, зачастую применяемых для оценки качества работы алгоритмов машинного обучения, стал набор данных под названием «The 20 Newsgroups data set» [11]. Представленный набор данных содержит разделенные на 20 классов пользовательские тексты из обезличенной переписки между людьми, что

позволяет провести аналогию с взаимодействием пользователя с персональным виртуальным ассистентом. Выбранный набор данных является независимым от данных бизнес-заказчика и позволяет провести первичную оценку корректности и адекватности работы алгоритмического обеспечения до создания набора промышленных данных и получения их от бизнес-заказчика. Предварительная оценка качества для анализа метрик производилась с применением алгоритмического обеспечения, используемого в готовом для внедрения в систему автоматизированной обработки запросов прототипе персонального виртуального ассистента.

В качестве метрик для анализа качества работы прототипа использовались метрики точность, полнота и F1-мера. Эксперимент по оценке качества алгоритмического обеспечения проводился по схеме перекрестной проверки с разбиением набора данных на 5 непересекающихся частей. В таблице 1 представлены результаты классификации синтетических данных «The 20 Newsgroups data set».

ТАБЛИЦА 1. РЕЗУЛЬТАТЫ КЛАССИФИКАЦИИ ДАННЫХ

Класс	Точность	Полнота	F1-мера	Размер данных
alt.atheism	0.752	0.734	0.743	319
comp.graphics	0.680	0.781	0.727	389
comp.os.ms-windows.misc	0.740	0.731	0.736	394
comp.sys.ibm.pc.hardware	0.716	0.735	0.725	392
comp.sys.mac.hardware	0.828	0.816	0.822	385
comp.windows.x	0.845	0.729	0.783	395
misc.forsale	0.764	0.897	0.825	390
rec.autos	0.905	0.891	0.898	396
rec.motorcycles	0.940	0.945	0.942	398
rec.sport.baseball	0.881	0.932	0.906	397
rec.sport.hockey	0.947	0.945	0.946	399
sci.crypt	0.931	0.889	0.910	396
sci.electronics	0.764	0.784	0.774	393
sci.med	0.895	0.838	0.866	396
sci.space	0.898	0.919	0.908	394
soc.religion.christian	0.788	0.925	0.851	398
talk.politics.guns	0.711	0.901	0.795	364
talk.politics.mideast	0.964	0.854	0.906	376
talk.politics.misc	0.780	0.594	0.674	310
talk.religion.misc	0.814	0.470	0.596	251
Доля правильных ответов	0.825	0.825	0.825	–
Среднее	0.827	0.815	0.817	7532
Среднее взвешенное	0.829	0.825	0.823	7532

В результате анализа метрик можно сделать окончательный вывод о значительной точности работы алгоритмического обеспечения и его возможности обобщать применение на различных наборах данных.

5. ЗАКЛЮЧЕНИЕ

В результате выполнения работы описан процесс создания алгоритмического обеспечения для функционирования персонального виртуального ассистента. Получены результаты оценки качества классификации пользовательских запросов на синтетических данных, а также получено время работы моделей машинного обучения. Усредненное взвешенное значение F1-метрики при работе двух моделей машинного обучения составило 82.3% при времени работы моделей логистической регрессии в 189 мс и модели fastText в 214 мс соответственно. Результаты свидетельствуют о применимости выбранных моделей в промышленной среде.

ЛИТЕРАТУРА

- [1] Maher, T. History of Speech & Voice Recognition and Transcription Software / T. Maher // Dragon Systems. – Centaur Academic Media DBA Centaur Multimedia, 2003.
- [2] Mishra, A.N. Business value appropriation roadmap for artificial intelligence / A.N. Mishra, A.K. Pani // VINE Journal of Information and Knowledge Management Systems. – 2020. – P. 353-368. DOI: 10.1108/VJKMS-07-2019-0107.
- [3] Zumalt, J.R. Voice recognition technology: Has it come of age? / J.R. Zumalt // Information technology and libraries. – 2005. – Vol. 24(4). – P. 180-185. DOI: 10.6017/ital.v24i4.3382.
- [4] Mahajan, R.S. Machine learning based paraphrase identification system using lexical syntactic features / R.S. Mahajan, M.A. Zaveri // IEEE International Conference on Computational Intelligence and Computing Research (ICIC). – 2016. – P. 1-5. DOI: 10.1109/ICIC.2016.7919721.
- [5] Khan, H.U. Using Machine Learning Techniques for Subjectivity Analysis based on Lexical and Nonlexical Features / H.U. Khan, A. Daud // International Arab Journal of Information Technology (IAJIT). – 2017. – Vol. 14(4).
- [6] LaValley, M.P. Logistic regression / M.P. LaValley // Circulation. – 2008. – Vol. 117(18). – P. 2395-2399. DOI: 10.1161/CIRCULATIONAHA.106.682658
- [7] fastText [Electronic resource]. — Access mode: <https://fasttext.cc/> (20.02.2022).
- [8] Pratiwi, N.I. Hate speech detection on Indonesian Instagram comments using FastText approach / N.I. Pratiwi, I. Budi, I. Alfina // International Conference on Advanced Computer Science and Information Systems (ICACSIS). – 2018. – P. 447-450. DOI: 10.1109/ICACSIS.2018.8618182.
- [9] Tiun, S. Classification of functional and non-functional requirement in software requirement using Word2vec and fast Text / S. Tiun, U.A. Mokhtar, S.H. Bakar, S. Saad // Journal of Physics: Conference Series. – 2020. – Vol. 1529(4). DOI: 10.1088/1742-6596/1529/4/042077.
- [10] Welcome to Python.org [Electronic resource]. — Access mode: <https://www.python.org/> (20.02.2022).
- [11] Home Page for 20 Newsgroups Data Set [Electronic resource]. — Access mode: <http://qwone.com/~jason/20Newsgroups/> (20.02.2022).

Scalogram-EMD distance for mobile ECGs

V. Guryanova
Lomonosov Moscow State University
Moscow, Russia
valeriiaguryanova@gmail.com

Abstract—Now there are devices that are capable of recording ECGs. The distance between signals can be helpful in classification problems for finding ECGs like the given one to know the expected disease scenario. This paper proposes a new distance based on the wavelet decomposition of the signal and earth mover's distance with a new base distance function. It is shown that the introduced distance is a metric over the considered signal equivalence classes. In addition, a method for creating new signals based on the developed distance has been proposed, which can be used to augment data when training deep neural networks. Finally, an experimental study has demonstrated that the generated signals can improve classification quality.

Keywords— ECG, Deep Learning, Augmentation, EMD, Wavelets.

1. INTRODUCTION

Now there are devices that can record ECGs (electrocardiogram). Examples of such devices are AliveCor [1], CardioQvark [2].

The distance between such signals can be a helpful feature. For example, one of these tasks may be searching for an ECG signal like chosen to find out the expected scenario for a patient's disease case, comparing it with patients treated in the past. Also, this metric can be used together with the k nearest neighbors (knn) [3] method and other classification methods that use distance.

Existing distances have drawbacks: many are not theoretical metrics, for example, the dynamic time warping algorithm [4]. Others do not consider the time component of the ECG signal. For example, such a metric was proposed in the paper [5] for stationary signals based on the Fourier transform and the Wasserstein distance (WF distance). The ECG signal is not stationary, as it depends on the electrical impulses coming from the heart at any given time. Therefore, the distance based on the Fourier transformation is not entirely applicable for such signals.

The paper is organized as follows. Chapter two defines distance and gives a theorem about its being a metric and some facts about its efficient calculation. Chapter three discusses new signals generation based on the created distance. Chapter four is devoted to the experimental study of distance compared to WF and the new augmentation technique. In conclusion, findings from the research are presented.

2. SCALOGRAM-EMD DISTANCE

A. Distance Definition

In this paper, it is proposed to use a metric based on the Wavelet decomposition of the signal since the wavelet decomposition has the property that its coefficients allow to analyze frequencies distributed over time [6]. The signal is represented as its normalized scalogram:

$$I_x(t, a) = \frac{|W_a(t)|^2}{\sum_t \sum_a |W_a(t)|^2}, \quad (1)$$

where W is the detail coefficients of wavelet decomposition, t is the time component of the wavelet decomposition, a is the scale component of the wavelet decomposition. The classical discrete wavelet transform was not used, but the stationary one [7] because of invariance to the time shift property.

Representing a signal as its normalized scalogram introduces an equivalence space $[x]=\{x'|I_{x'}(t,a)=I_x(t,a)\}$ on the set of signals. Thus, signals are considered equivalent if one is obtained from the other by multiplying by a coefficient. If not all levels are considered, the signals are considered to be equivalent to a certain frequency level contained in the first or last coefficients. Thus, the distance will be considered not over the signals themselves but the equivalence classes of these signals. In many machine learning problems can be assumed that signals that fall into the same equivalence class belong to the same class.

The Earth Movers Distance metric [8] with the base metric $L_{1\alpha}$ is used to determine the distance.

- Let the matrices P and Q have m rows and n columns and $N = m \times n$;
- The list of indexes for values is defined as $\mathcal{J} = \{(i, j): 1 \leq i \leq m, 1 \leq j \leq n\}$;
- The list of indexes for threads is defined as $\mathcal{J} = \{(i, j, k, l): (i, j) \in \mathcal{J}, (k, l) \in \mathcal{J}\}$;
- Matrices $P = \{p_{i,j}: (i, j) \in \mathcal{J}\}$ and $Q = \{q_{i,j}: (i, j) \in \mathcal{J}\}$ - matrices between which the distance is being searched.

$$EMD(P, Q) = \min_{F=\{f_{i,j,k,l}: (i,j,k,l) \in \mathcal{J}\}} \sum_j f_{i,j,k,l} d_{i,j,k,l}$$

$$\begin{cases} \sum_{(k,l) \in \mathcal{J}} f_{i,j,k,l} = p_{i,j} & \forall (i, j) \in \mathcal{J} \\ \sum_{(i,j) \in \mathcal{J}} f_{i,j,k,l} = q_{k,l} & \forall (k, l) \in \mathcal{J} \\ f_{i,j,k,l} \geq 0 & \forall (i, j, k, l) \in \mathcal{J} \end{cases}$$

where F is the flow from P to Q , and $f_{i,j,k,l}$ defines the flow from (i, j) to (k, l) .

In this paper, it is proposed to define the base distance as:

$$L_{1\alpha}: d_{i,j,k,l} = |i - k| + \alpha|j - l|, \quad \alpha \in (0,1].$$

This distance defines the distance between two points as a successive transfer in frequency and time. At the same time, it is proposed to introduce an additional coefficient $\alpha \in (0,1]$, which will show how much the transfer in time is less significant than the transfer in frequency.

Definition 1. Scalogram-EMD (SE) distance between two signal equivalence classes $[x]$ and $[y]$ is defined as:

$$SE([x], [y]) = EMD_{L_{1\alpha}}(I_x(t, a), I_y(t, a)).$$

Theorem 1. SE is a metric over signal equivalence classes.

Б. Efficient way to calculate $EMD_{L_{1\alpha}}$.

The article [9] proposes a method for quickly and accurately finding $EMD_{L_{1\alpha}}$ in the case of $\alpha = 1$. The number of unknowns in the equations for finding the minimum flow was reduced to $O(N)$ from $O(N^2)$ in matrices consisting of N points. Empirically, it was shown that the new algorithm has $O(N^2)$ time, while the classical algorithm shows $O(\log n N^3)$ complexity.

In this work, it was shown that the algorithm constructed for L_1 could be adapted to find $L_{1\alpha}$ without increasing the level of asymptotic complexity.

3. NEW SIGNALS GENERATION WITH SCALOGRAM-EMD DISTANCE

The additional synthetic data can be used in the online and offline augmentation approaches to improve the model's quality on the original test cases. This paper proposes a new method for generating data, which uses the flow obtained by calculating the distance between two signals.

Two signals belonging to the same class are selected from the training set. Two matrices similar to formula (1) are built for each signal. One is based on the approximation coefficients, the other is based on the detail coefficients. In this case, both coefficients are needed for the inverse transformation from the wavelet decomposition to the signal. The matrices correspond to the signs of the approximation and detail matrices stored before they are squared.

Then a flow is built based on the EMD distance between each pair of matrices. The main idea of new signal generation is to apply only a part of the transformations from the entire flow that translates one signal into another. For example, suppose we use a random part of the flow from EMD distance. In that case, the resulting matrix may no longer correspond to the original class of matrices (it may contain negative values, and the sum of all its elements may no longer be equal to one).

The resulting flow is built according to the following principles. 1) For the resulting flow, single flows are selected from those points where there are no incoming flows. 2) As long as there are single flows not included in the resulting flow: points are selected to which there are no incoming flows, except for those in the resulting flow; flows from these points are added to the resulting flow.

Theorem 2. The flow built according to the above principle allows keeping the non-negativity of matrices and the sum of their elements equal to 1 when applying any starting part of the resulting flow.

Part of the transformed flow is applied to the original matrices. The root is extracted from the new matrices. The sign is determined as the sign of the closest original matrix in the point. The inverse wavelet transform is performed. Below is an example of the newly generated signal.

TABLE I. DATASET

Dataset	Experiments		
	Metric Name	WF	SE
Arithmia Detection	f-weighted	0.51	0.55
Tuberculosis Patient	F1-score	0.32	0.38
Patient Detection	Patient Detection	0.01	0.06

4. EXPERIMENTS AND RESULTS

Experiments were carried out on three different datasets: dataset for the presence of tuberculosis, identification of the patient by his ECG, and determination of arrhythmia type. For the first task, the f1 score was used. For the second task, accuracy was used, and for the last task f-weighted score was used. The knn method was used as a classifier. In the case of an extensive training sample, the WF distance selected 383 candidates. SE was used to choose the closest ones from them.

During the experiments, it was found that changing the α parameter positively affects improving the quality.

Also, synthetically generated signals from paragraph III were added to the training data from the article [10]. The same neural network and test set were used. The quality was improved from 0.934 to 0.952 ROC AUC.

5. CONCLUSION

On the selected test problems, the classification quality of the method using the WE distance shows a higher performance than the method using the WF distance. The α parameter improves the classification quality and can be used for tuning. The generated signals using the EMD distance improve the classification quality on the considered problem.

REFERENCES

- [1] AliveCor Web Site [Electronic resource]. – Access mode: <https://www.alivecor.com> (20.01.2022).
- [2] CardioQvark Web Site [Electronic resource]. – Access mode: <https://www.cardioqvark.com> (20.01.2022).
- [3] Dangeti, P. Statistics for machine learning / P. Dangeti. – Packt Publishing Ltd, 2017.
- [4] Müller, M. Dynamic time warping / M. Müller // Information retrieval for music and motion. – 2007. – P. 69-84.
- [5] Cazelles, E. The Wasserstein-Fourier distance for stationary time series / E. Cazelles, A. Robert, F. Tobar // IEEE Transactions on Signal Processing. – 2020. – Vol. 69. – P. 709-721.
- [6] Chun-Lin, L. A tutorial of the wavelet transform / L. Chun-Lin // NTUEE. – 2010. – Vol. 21. – P. 22.
- [7] Nason, G.P. The stationary wavelet transform and some statistical applications / G.P. Nason, B.W. Silverman // Wavelets and statistics. – Springer, New York, NY, 1995. – P. 281-299.
- [8] Rubner, Y. The earth mover's distance as a metric for image retrieval / Y. Rubner, C. Tomasi, L.J. Guibas // International journal of computer vision. – 2000. – Vol. 40(2). – P. 99-121.
- [9] Ling, H. An efficient earth mover's distance algorithm for robust histogram comparison / H. Ling, K. Okada // IEEE transactions on pattern analysis and machine intelligence. – 2007. – Vol. 29(5). – P. 840-853.
- [10] Guryanova, V. Transfer Learning for tuberculosis screening by single-channel ECG / V. Guryanova // International Conference on Information Technology and Nanotechnology (ITNT-2020). – 2020. – P. 129-134.

Oblivious piecewise-linear decision trees

A. Gurianov
Lomonosov Moscow State University
Moscow, Russia
guryanov93@gmail.com

Abstract—Gradient boosting ensembles of decision trees are a very popular type of machine learning model, with several popular implementations. Some of those implementations utilize symmetric decision trees - decision trees of a specific structure that improve regularization and speed predictions. Over the last years, several research works have been published related to piecewise-linear decision trees and their utilization in gradient boosting ensembles. In this paper, symmetric piecewise-linear decision trees were introduced, and it was shown that it is possible to efficiently train such trees in gradient boosting ensembles. It was shown that such symmetric piecewise-linear decision trees have significantly faster prediction time compared to regular decision trees and piecewise-linear decision trees of similar depths and that ensembles of symmetric piecewise-linear decision trees achieve competitive quality on open datasets.

Keywords— machine learning, decision trees, piecewise-linear decision trees, gradient boosting.

I. INTRODUCTION

Gradient boosting over decision trees is one of the most popular machine learning algorithms. It was initially proposed in [1], and it can achieve excellent quality on a large variety of datasets out of the box, has built-in feature selection and regularization, and also has several efficient implementations [2], [3], [4]. Improving efficiency and performance of gradient boosting ensembles are popular areas of scientific research. One of the possible improvements to those algorithms is the usage of oblivious (or symmetric) decision trees [5] that have the same split rule on each level. This decision tree structure drastically reduces prediction time of a single decision tree, which is important in many big data and highload scenarios. Another possible type of tree that possesses interesting properties is a piecewise-linear decision tree. Piecewise-linear decision trees return a linear function over features in a leaf as a prediction. This allowing to more easily model linear dependencies in the data and improves performance without increasing prediction times.

In this paper, a definition of oblivious piecewise-linear decision trees is proposed. Algorithms for training and predicting such trees were implemented using Python, and a significant boost in prediction speed is demonstrated on publicly available datasets.

This paper is organized as follows. First, a brief overview of gradient boosting over decision trees is given. Then the concept of piecewise-linear decision trees is explained. Then a definition of oblivious piecewise-linear decision tree is given, and an algorithm for efficient training of such trees within gradient boosting paradigm is described. The next section describes an experimental evaluation of prediction speed of suggested algorithms. The last chapter contains the conclusion of the work.

II. GRADIENT BOOSTING OVER DECISION TREES

Given a dataset of samples $D = (x_i, y_i)$ with m features, where $|D| = n$ is the number of samples, and $x_i \in R^m$, gradient boosting algorithm trains a sequence of size T of

decision trees $\{t_k\}_1^T$ to minimise global loss function L . The final output of gradient boosting is the summation of these trees $\hat{y} = \sum_{k=1}^T t_k(x_i)$. Let $l: R^2 \rightarrow R$ be the loss function for a single sample. Then the global loss function is usually represented as a summation of individual losses over training dataset D and a regularization term $\Omega(t_k)$ to prevent overfitting:

$$L = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^T \Omega(t_k).$$

Let $\hat{y}_i^{(k)}$ be the predicted value of x_i after iteration k . At iteration $k + 1$ a new tree t_{k+1} is trained to minimise the following loss

$$\begin{aligned} L^{(k+1)} &= \sum_{i=1}^n l(\hat{y}_i^{(k+1)}, y_i) + \sum_{k=1}^T \Omega(t_k) \\ &= \sum_{i=1}^n l(\hat{y}_i^{(k)} + t_k(x_i), y_i) + \sum_{k=1}^T \Omega(t_k). \end{aligned}$$

With second-order approximation, and assuming all predictors from 1 to k fixed, the loss at step k can be represented as

$$L^{(k+1)} \approx C + \Omega(t_k) + \sum_{i=1}^n (g_i t_k(x_i) + \frac{1}{2} h_i t_k(x_i)^2),$$

Popular libraries for gradient boosting over decision trees implement algorithms that minimize global loss in the way described above. Piecewise-linear decision trees described in [6] also fit linear coefficients for features in the split. For such trees it is possible to represent loss component of a leaf parametrized with linear coefficient w and constant value b the following way:

$$\begin{aligned} GHF_N &= \sum_{i \in N} g_i + h_i f_i, GXHF_N^j = \sum_{i \in N} g_i x_i^j + h_i f_i x_i^j \\ H_N &= \sum_{i \in N} h_i, HX_N^j = \sum_{i \in N} h_i x_i^j, HXX_N^j = \sum_{i \in N} h_i (x_i^j)^2. \\ L_{leaf}^j &= \frac{1}{2} w^2 (HXX_N^j + \lambda_w) + \frac{1}{2} b^2 (H_N + \lambda_b) \\ &\quad + w \cdot GXHF_N^j + b \cdot GHF_N \\ &\quad + wb \cdot HX_N^j + \sum (g_i f_i + \frac{1}{2} h_i f_i^2). \end{aligned}$$

III. OBLIVIOUS PIECEWISE-LINEAR DECISION TREES

A. Definition

Definition 1 (Symmetric piecewise-linear decision tree). A symmetric piecewise-linear decision tree is a piecewise-linear decision tree that has all the levels of the tree completely filled, all nodes on the same level have the same splitting rule and common left linear coefficient and right linear coefficient.

Such kind of trees are simpler than general piecewise-linear decision trees and can be considered as a method of regularization. They also have an increased prediction speed, having a greater degree of utilization of modern CPUs through SIMD parallelization.

B. Efficient Learning

Algorithm for efficient learning of oblivious piecewise-linear decision trees has a lot in common with algorithm described in [6], with several key differences. The general idea remains the same - at each step of the algorithm we choose to change the tree structure to optimize global loss function based on values of gradients and Hessians in each training instance computed from gradient descent step.

First difference is that tree building is performed level-wise instead of node-wise. Second difference is that we build each level of the tree by considering every possible split over every possible feature that optimizes global loss function assuming common splitting rules and left and right linear coefficients for nodes on the same level.

$$L_{level}^j = \sum_{l \in Leaves} \left(\frac{1}{2} w^2 (HXX_N^j + \lambda_w) + \frac{1}{2} b_l^2 (H_N + \lambda_{b_l}) + w \cdot GXHFX_N^j + b_l \cdot GHF_N + w \cdot b_l \cdot HX_N^j \right).$$

Finding optimal values of linear coefficients, leaf values on each level and optimal value of loss function corresponding to each splitting rule is a quadratic form minimization problem. The matrix representation of this quadratic form is an arrowhead matrix, which has efficient algorithms for computing an inverse matrix [7]. This inverse algorithm is used to solve a minimization problem and find optimal values of loss, linear coefficients, and leaf values at each level. Building of levels in the oblivious piecewise-linear decision tree continues until stop condition is achieved, usually formulated as a limit on the number of levels.

IV. EXPERIMENTS

Suggested algorithm for learning and predicting gradient boosting ensembles of oblivious piecewise-linear decision trees was implemented using Python programming language. To demonstrate improvements in prediction performance, ensembles of 100 decision trees with depth 5 were trained on a HIGGS dataset that consists of 11000000 objects with 28 real-valued features using XGBoost, CatBoost, Piecewise-Linear decision trees (PWL) and Oblivious Piecewise-Linear decision trees (SYMPWL). These ensembles were used to predict 5% of HIGGS dataset 100 times, and averages prediction times were reported in the table below. According to the results, oblivious piecewise-linear decision trees achieved a significant improvement in prediction time compared to regular piecewise-linear decision trees but are slower than CatBoost oblivious decision trees.

TABLE I. HIGGS DATASET FEATURES USING XGBOOST, CATBOOST, PWL AND SYMPWL

Library	Prediction time (ms)
XGBoost	280
CatBoost	90
PWL	300
SYMPWL	132

V. CONCLUSION

Efficient machine learning algorithms that are able to demonstrate the best quality possible in as little prediction time as possible is an important area of research relevant for many industries. This paper introduces oblivious piecewise-linear decision trees - a type of decision tree that combines advantages of oblivious decision trees and piecewise-linear decision trees. Suggested oblivious piecewise-linear decision tree is possible to efficiently train and it is possible to implement an algorithm that achieves a significant prediction speedup per tree compared to XGBoost and regular piecewise-linear decision trees.

REFERENCES

- [1] Friedman, J.H. Stochastic gradient boosting / J.H. Friedman // Computational statistics & data analysis. – 2002. – Vol. 38(4). – P. 367-378.
- [2] Ke, G. Lightgbm: A highly efficient gradient boosting decision tree / G. Ke // Advances in neural information processing systems. – 2017. – Vol. 30.
- [3] Chen, T. Xgboost: A scalable tree boosting system / T. Chen, C. Guestrin // Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. – 2016. – P. 785-794.
- [4] Dorogush, A.V. CatBoost: gradient boosting with categorical features support / A.V. Dorogush, V. Ershov, A. Gulin // ArXiv preprint: 1810.11363, 2018.
- [5] Kohavi, R. Oblivious decision trees, graphs, and top-down pruning / R. Kohavi, C.H. Li // IJCAI. – 1995. – P. 1071-1079.
- [6] Guryanov, A. Histogram-based algorithm for building gradient boosting ensembles of piecewise linear decision trees / A. Guryanov // International Conference on Analysis of Images, Social Networks and Texts. – Springer, Cham, 2019. – P. 39-50.
- [7] Najafi, H.S. An efficient method for computing the inverse of arrowhead matrices / H.S. Najafi, S.A. Edalatpanah, G.A. Gravvanis // Applied Mathematics Letters. – 2014. – Vol. 33. – P. 1-5.

Платформа имитационного моделирования оборудования инфраструктурных объектов в распределенной среде

И.В. Бычков
Институт динамики систем и
теории управления им. В.М.
Матросова СО РАН
Иркутск, Россия
idstu@icc.ru

А.Г. Феоктистов
Институт динамики систем и
теории управления им. В.М.
Матросова СО РАН
Иркутск, Россия
agf@icc.ru

Р.О. Костромин
Институт динамики систем и
теории управления им. В.М.
Матросова СО РАН
Иркутск, Россия
kostromin@icc.ru

Аннотация—В докладе предлагается новая платформа на базе микросервисного подхода к организации имитационного моделирования в гетерогенных распределенных вычислительных средах. В рамках предлагаемого подхода все операции, связанные с подготовкой данных, выполнением моделей и анализом полученных результатов, выполняются микросервисами. Основными преимуществами предлагаемого подхода являются многовариантные вычисления в рамках имитационного моделирования и возможность интеграции ресурсов суперкомпьютерных центров общего доступа с облачными и туманными платформами. Кроме того, обеспечивается автоматизированное создание веб-форм микросервисов с использованием спецификаций моделей. Данная платформа применяется для изучения экологически чистого оборудования объектов Байкальской природной территории, к которому относятся тепловые насосы. В рамках исследования проведена оценка затрат на возможное использование тепловых насосов в различных эксплуатационных и метеорологических условиях для типового объекта. Проведенный сравнительный анализ подтвердил преимущества применения предложенного подхода.

Ключевые слова— имитационное моделирование, микросервисы, распределенные вычисления, инфраструктурные объекты, метеорологические данные.

1. ВВЕДЕНИЕ

В настоящее время одной из глобальных проблем является бережное и эффективное использование природных ресурсов. В частности, практически значимой проблемой является снижение энергопотребления. Таким образом, возникает необходимость в проведении исследований по моделированию, прогнозированию и контролю эффективности использования ресурсов при эксплуатации различных объектов, связанных с деятельностью человека. В таких исследованиях требуется использование данных как об изучаемых объектах, так и об окружающей среде. В частности, важную роль здесь играет сбор и анализ метеорологических данных (облачности, солнечная радиация, температуры и др.), поскольку такие атмосферные параметры могут существенно повлиять на энергопотребление.

2. ЦИФРОВЫЕ ДВОЙНИКИ СЛОЖНЫХ СИСТЕМ

Современное состояние развития вычислительных и телекоммуникационных технологий дает возможность быстро получать, оцифровывать и эффективно анализировать большие данные. Часто такие данные

генерируются в процессах функционирования или моделирования сложных систем в различных сферах жизнедеятельности человека. Новой формой изучения подобных систем, основанной на использовании текущих и ретроспективных данных, является так называемый цифровой двойник (ЦД) [1].

Как правило, ЦД – это виртуальная программная сущность, которая отражает наиболее важные компоненты изучаемого объекта в его жизненном цикле. В рамках ЦД применяются доступные физические и виртуальные данные об объекте. Основным инструментом исследования объекта с помощью ЦД по-прежнему является имитационное моделирование [2].

Одним из важных направлений использования ЦД является поддержка принятия решений при управлении социально-экономическими территориями и расположенными на них объектами инфраструктуры [3]. При этом упор делается на использование экологически чистых технологий и оборудования. Как правило, компоненты оборудования размещаются в распределенной среде и контролируются различными субъектами, представляющими объект.

Имитационное моделирование – перспективное направление внедрения ЦД для исследования инфраструктурных объектов. Оно широко используется как один из основных инструментов поддержки принятия решений при управлении различными техническими, экономическими, культурными, социальными и другими системами [4]. В рамках такого управления решаются широкий круг практических задач. Они включают принятие стратегических решений, планирование, контроль, улучшение бизнес-процессов и внедрение новых технологий, в том числе экологически чистых технологий. Имитационное моделирование позволяет получать информацию о выполнении бизнес-процессов в различных инфраструктурах с различной конфигурацией.

Анализ результатов моделирования работы объекта позволяет существенно улучшить его основные параметры. К сожалению, в современных средствах моделирования известен ряд проблем [5]. Среди них сложность разработки моделей, их параллельное выполнение, поддержка удобного доступа пользователей для подготовки и проведения масштабных экспериментов и др. Более того, обработка больших данных в гетерогенной среде по-прежнему остается проблемой.

Упомянутые проблема относятся к таким распространенным системам имитационного

моделирования, таких как GPSS [6] и AnyLogic [7]. Они позволяют с необходимой степенью детализации описать предметную область и выполнить моделирование. Специализированные подсистемы позволяют дополнительно отслеживать изменения переменных, строить таблицы и графики. Однако в GPSS эти подсистемы имеют ограниченные функциональные возможности, а AnyLogic и подобные ей системы являются коммерческими и их применение в научно-исследовательской деятельности не всегда оправдано. При этом потребность в различных средствах визуализации в процессе имитационного моделирования остается актуальной [8].

Таким образом, исследование направлено на разработку новых сервис-ориентированных инструментов для подготовки и проведения масштабных научных экспериментов с имитационными моделями для инфраструктурных объектов на примере Байкальской природной территории. Важную роль играют подсистемы выполнения имитационных моделей в распределенной вычислительной среде. В качестве основы для выполнения моделей используется GPSS.

3. ПЛАТФОРМА ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Ключевыми особенностями новой платформы являются: возможность автоматизированной подготовки вычислительной среды (запуск виртуальных машин, доставка модулей и входных данных); генерация веб-интерфейса для работы с моделью по ее спецификации; запуск множества экземпляров GPSS в распределенной среде с разными наборами входных данных в рамках многовариантных вычислений; визуализация результатов моделирования и их анализ [9].

В работе проведен сравнительный анализ функциональных возможностей для следующих систем: GPSS, AnyLogic и разработанной платформы. Опираясь на опыт пользователей, определены наиболее важные возможности этих систем. Среди них распределенные вычисления, однокритериальная оптимизация, многокритериальный выбор, гибридное моделирование с применением агентов, мульти-облачные вычисления и поддержка типовых рабочих процессов (workflow).

В качестве практической апробации новой платформы проведен вычислительный эксперимент применительно к инфраструктурным объектам Байкальской природной территории. В рамках эксперимента реализована методика оценки экономической эффективности экологически чистого теплоснабжения типового оздоровительного туристического центра. одной из практических целей моделирования возможной замены источника тепла является оценка уменьшения выбросов CO₂ от сжигания дизельного топлива или угля. Еще один немаловажный фактор – это стоимость теплового насосного оборудования. Полученные результаты могут быть использованы в качестве рекомендаций для специалистов по эксплуатации оборудования. Для обеспечения имитационной модели релевантными метеорологическими данными реализован сбор текущих метеорологических показателей от датчиков, работа которых согласуется с концепцией туманных вычислений.

4. ЗАКЛЮЧЕНИЕ

Разработаны сервис-ориентированные средства для поддержки высокопроизводительных вычислений,

однокритериальной оптимизации, многокритериального выбора, гибридного имитационного моделирования с применением агентов и мульти-облачных вычислений. Эти инструменты позволяют проводить имитационное моделирование в гетерогенных распределенных средах с использованием GPSS-моделей.

На основе разработанного инструментария созданы сервисы оценки экономической эффективности типовой туристической базы отдыха, расположенной на Байкальской природной территории. С помощью сервисов проведена оценка стоимости применения экологически чистой технологии отопления тепловыми насосами в различных эксплуатационных и метеорологических условиях.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке РФФИ и Правительства Иркутской области, проект № 20-47-380002-р_а «Математическое и информационное моделирование инфраструктурных объектов Байкальской природной территории», а также Министерства науки и высшего образования Российской Федерации, проект № FWEW-2021-0005 «Технологии разработки и анализа предметно-ориентированных интеллектуальных систем группового управления в недетерминированных распределенных средах».

ЛИТЕРАТУРА

- [1] Петров, А.В. Имитация как основа технологии цифровых двойников / А.В. Петров // Вестник Иркутского государственного технического университета. – 2018. – Т. 10, № 141. – С. 56-66. DOI: 10.21285/1814-3520-2018-10-56-66.
- [2] Ильясов, Б.Г. Методологические основы моделирования и интеллектуального управления промышленным комплексом как сложным динамическим многоагентным объектом / Б.Г. Ильясов, Е.А. Макарова, Е.Ш. Закиева, Э.Р. Габдуллина // Современные наукоемкие технологии. – 2020. – Т. 11, № 2. – С. 288-293. DOI: 10.17513/snt.38376.
- [3] Медведев, А.В. Цифровые двойники территорий для поддержки принятия решений в сфере регионального социально-экономического развития / А.В. Медведев // Современные наукоемкие технологии. – 2020. – Т. 6, № 1. – С. 61-66. DOI: 10.17513/snt.38072.
- [4] Феоктистов, А.Г. Цифровые двойники процессов работы природосберегающего оборудования инфраструктурного объекта / А.Г. Феоктистов, Р.О. Костромин, И.А. Сидоров, С.А. Горский, О.Ю. Башарина // Современные наукоемкие технологии. – 2021. – № 1. – С. 57-62.
- [5] Kostromin, R. Microservice-Based Approach to Simulating Environmentally-Friendly Equipment of Infrastructure Objects Taking into Account Meteorological Data / R. Kostromin, O. Basharina, A. Feoktistov, I. Sidorov // Atmosphere. – 2021. – Vol. 12(9). – P. 1-24.
- [6] Schriber, T.J. Simulation Using GPSS. – John Wiley & Sons, 1974. – 533 p.
- [7] AnyLogic [Электронный ресурс]. – Режим доступа: <https://www.anylogic.com> (10.03.2022).
- [8] Pliatsios, D. A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics / D. Pliatsios, P. Sarigiannidis, T. Lagkas, A.G. Sarigiannidis // IEEE Communications Surveys & Tutorials. – 2020. – Vol. 22(3). – P. 1942-1976. DOI: 10.1109/COMST.2020.2987688.
- [9] Феоктистов, А.Г. Сервис-ориентированный подход к имитационному моделированию процессов функционирования инфраструктурных объектов / А.Г. Феоктистов, Р.О. Костромин, И.А. Сидоров // Современные и наукоемкие технологии. – 2020. – № 9. – С. 76-81.



Министерство образования
и науки Самарской области



САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

Самарский университет

ИСОИ  **ИПСИ**

Институт систем обработки изображений РАН



Institute of Electrical
and Electronics Engineers



Photronics



ФОТОНИКА



Компьютерная оптика