

РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ИЗМЕНЕНИЯ СТОИМОСТИ ЖИЛЬЯ

Булгаков А.С.
charlez.wape@gmail.com

Голова А.А.
nastyagolova13@gmail.com

Сопченко Е.В.
viljevna@ya.ru

Введение

Индустрия недвижимости — динамичный бизнес, влияющий на каждый сектор жизни. По определению, отрасль включает все аспекты владения землей и зданиями. Качество застроенной среды влияет на качество повседневной жизни. Искусственная нейронная сеть (ИНС) (англ. Artificial neural network (ANN)) — упрощенная модель биологической нейронной сети, представляющая собой совокупность искусственных нейронов, взаимодействующих между собой.

Постановка задачи и выбор методов решения

Цель исследования: разработать автоматизированную систему прогнозирования изменения стоимости жилья с помощью искусственной нейронной сети и получить наиболее оптимальные параметры для нее.

Во время разработки системы выбор структуры сети для решения задачи прогнозирования изменения стоимости жилья проводился среди следующих сетей:

- многослойный персептрон;
- радиально-базисные сети;
- обобщенно-регрессионные сети;
- сети Вольтерри;
- сети Эльмана.

Среди вышеперечисленных видов сетей была выделена радиально-базисная нейронная сеть. Радиально-базисные сети имеют двухслойную структуру. Нелинейное преобразование входных сигналов происходит в пределах скрытого слоя нейронов, имеющих радиальные функции активации. В качестве такой радиальной функции является функция Гаусса:

$$\varphi(x) = \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right)$$

где x — входные векторы; c_j — центры; σ_j — радиусы, определяющие величину области, в которой значение функции будет отличным от нуля.

Выбор алгоритма обучения нейронной сети

В случае решения задачи прогнозирования стоимости жилья был выбран метод обучения из категории с учителем, так как специфика задачи подразумевает наличие ожидаемого результата при обучении. Таким образом, при разработке сравнивались метод самоорганизации и обратного распространения ошибки.

Параметры оценки качества прогноза

В случае прогнозирования стоимости жилья, выборку данных можно разделить на части: для обучения и для тестирования. Таким образом, часть данных, предназначенная для тестирования, может быть использована для проверки сети на работоспособность и оценки ее точности.

Для оценки качества прогноза при подборе оптимальных параметров использовались значения: среднеквадратического отклонения и коэффициента несовпадения Тейла.

Получение оптимальных значений нейронной сети

Для проведения подбора оптимальных параметров была использована реальная выборка стоимости жилья, значения в данной выборке были указаны с периодом в два месяца. Оптимальные значения подбирались для следующих параметров: число нейронов скрытого слоя, число входов сети, объем выборки, коэффициент обучения, число итераций. Данные полученные при подборе оптимальных значений приведены в таблицах:

Число нейронов	СКО		Коэффициент Тейла	
	Наискорейшего спуска	Самоорганизации	Наискорейшего спуска	Самоорганизации
5	0,21063	0,28865	0,22757	0,2589
10	0,12952	0,20882	0,20253	0,21844
15	0,07299	0,20869	0,18327	0,20869
20	0,06269	0,06982	0,18167	0,19651
25	0,03274	0,03408	0,18003	0,19306
30	0,09677	0,13672	0,19183	0,19892
35	0,11159	0,24188	0,19904	0,20188

Зависимость погрешности прогноза от числа нейронов скрытого слоя

Число входов	СКО		Коэффициент Тейла	
	Наискорейшего спуска	Самоорганизации	Наискорейшего спуска	Самоорганизации
5	0,03691	0,04066	0,06719	0,06806
10	0,02501	0,04316	0,10469	0,12098
15	0,09611	0,07075	0,13314	0,15949
20	0,03274	0,03408	0,11289	0,10365
25	0,06419	0,05271	0,15888	0,15616
30	0,06653	0,06961	0,16299	0,25396
35	0,07466	0,12995	0,24828	0,44113

Зависимость погрешности прогноза от числа входов сети

Объем выборки	СКО		Коэффициент Тейла	
	Наискорейшего спуска	Самоорганизации	Наискорейшего спуска	Самоорганизации
185	0,04869	0,04785	0,08103	0,07899
215	0,03939	0,05082	0,07615	0,09634
245	0,04502	0,04577	0,09152	0,08704
275	0,05134	0,03827	0,09890	0,07263
305	0,04314	0,04644	0,08302	0,08838
335	0,05269	0,03903	0,09914	0,07169
365	0,03691	0,04066	0,06719	0,06806

Зависимость погрешности от объема выборки

Результаты

В результате проведенного исследования на реальном наборе значений стоимости жилья были получены следующие значения параметров: количество входов сети 5, количество нейронов скрытого слоя 25, обучающая выборка состоит из 365 записей стоимости жилья.

