

# Разработка автоматизированной системы прогнозирования изменения стоимости жилья

Е.В. Сопченко  
Самарский национальный  
исследовательский университет  
Самара, Россия  
viljevna@ya.ru

А.С. Булгаков  
Самарский национальный  
исследовательский университет  
Самара, Россия  
charlez.wape@gmail.com

А.А. Голова  
Самарский национальный  
исследовательский университет  
Самара, Россия  
nastyagolova13@gmail.com

**Аннотация** — статья посвящена созданию автоматизированной системы прогнозирования изменения стоимости жилья при помощи радиально-базисной нейронной сети и поиску оптимальных параметров для нее. Ввиду активного расширения рынка недвижимости и наличия малого количества простых решений для прогнозирования, создание данной автоматизированной системы актуально. Результаты прогноза и сравнения структур нейронных сетей были получены на реальных наборах данных.

**Ключевые слова** — недвижимость, радиально-базисная нейронная сеть, анализ данных, стоимость, прогнозирование

## 1. 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 [1].

## 2. 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 [2].

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 (1):

$$\varphi(x) = \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right) \quad (1)$$

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

## 3. ВЫБОР АЛГОРИТМА ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Обучение радиально-базисной сети состоит в подборе таких весов нейронов выходного слоя  $w_j$  и параметров радиальных функций  $c_j$  и  $\sigma_j$ , которые отобразят сложные зависимости между данными, подаваемыми на вход сети, и выходными данными [3]. Таким образом, в процессе обучения сети необходимо минимизировать погрешность прогнозирования.

Алгоритмы обучения отличаются для разных структур сетей.

Можно выделить две основные категории алгоритмов:

- обучение с учителем;
- обучение без учителя [3].

Обучение с учителем подразумевает наличие обучающих пар вида: входное значение и ожидаемый результат [4]. В случае обучения без учителя нейронная сеть, подчиняясь некоторым внутренним процессам самоорганизации, начинает решать некоторую задачу.

В случае решения задачи прогнозирования стоимости жилья целесообразно выбрать 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.

Метод самоорганизации основан на разделении пространства обучающих данных на области Вороного. Данные внутри области группируются вокруг центральной точки, отождествляемой с центром радиальной функции, являющейся средним значением элементов области.

## 4. РЕАЛИЗАЦИЯ АЛГОРИТМА ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

В данной работе применяется накопительная версия алгоритма. В соответствии с ней все обучающие векторы подаются сразу. После этого на первом этапе множество векторов, приписанных одному центру, образуют кластер с центром, определяемым по формуле (2)

$$c_i(k+1) = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j(k), \quad (2)$$

где  $N_i$  – количество векторов  $x_j(k)$ ;  $x_j(k)$  – входные векторы, приписанные в  $k$ -ом цикле  $i$ -ому центру.

На втором этапе работы алгоритма при фиксированных значениях центров производится расчет радиусов. В данной работе используется версия метода, в которой радиус  $\sigma_i$  равняется евклидову расстоянию между центром  $c_i$  и его ближайшим соседом  $c_j$ .

Метод обратного распространения ошибки также состоит из двух этапов:

1) Рассчитываются значения дифференциалов функции ошибки по формулам (3) и (4) для каждой компоненты центра и радиуса функции каждого нейрона скрытого слоя:

$$\frac{dE}{dc_j} = (y_i - d_i)w_j e^{-\frac{(x_i - c_j)^2}{2\sigma_j^2}} \frac{x_i - c_j}{\sigma_j^2} \quad (3)$$

$$\frac{dE}{d\sigma_j} = (y_i - d_i)w_j e^{-\frac{(x_i - c_j)^2}{2\sigma_j^2}} \frac{(x_i - c_j)^2}{\sigma_j^3} \quad (4)$$

2) Уточняются компоненты центров и радиусов по формулам (5) и (6):

$$c_j(k+1) = c_j(k) - \eta \frac{dE}{dc_j} \quad (5)$$

$$\sigma_j(k+1) = \sigma_j(k) - \eta \frac{dE}{d\sigma_j} \quad (6)$$

После обучения нейронной сети необходимо оценить качество данной сети. 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.

#### 5. 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 1-3.

Таблица 1 Зависимость погрешности прогноза от числа нейронов скрытого слоя

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	<b>0,03274</b>	<b>0,03408</b>	<b>0,18003</b>	<b>0,19306</b>
30	0,09677	0,13672	0,19183	0,19892
35	0,11159	0,24188	0,19904	0,20188

Таблица 2 Зависимость погрешности прогноза от числа входов сети

Число входов	СКО		Коэффициент Тейла	
	Наискорейшего спуска	Самоорганизации	Наискорейшего спуска	Самоорганизации
5	<b>0,03691</b>	<b>0,04066</b>	<b>0,06719</b>	<b>0,06806</b>
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

Таблица 3 Зависимость погрешности от объема выборки

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
<b>365</b>	<b>0,03691</b>	<b>0,04066</b>	<b>0,06719</b>	<b>0,06806</b>

#### 6. ЗАКЛЮЧЕНИЕ

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.

#### ЛИТЕРАТУРА

- [1] Предметная область [Электронный ресурс]. – Режим доступа: [https://neerc.ifmo.ru/wiki/index.php?title=Нейронные\\_сети,\\_перцептрон](https://neerc.ifmo.ru/wiki/index.php?title=Нейронные_сети,_перцептрон) (10.11.2022).
- [2] Солдатова, О.П. Применение нейронных сетей для решения задач прогнозирования / О.П. Солдатова, В.В. Семенов – Самара: Издательство Самарского национального исследовательского университета имени С.П. Королева, 2018. – 344 с.
- [3] Осовский, С.О. Нейронные сети для обработки информации / С.О. Осовский; пер. с польск. И.Д. Рудинского, рец. И.Б. Фоминых. – Москва: Горячая линия - Телеком, 2016. – 448 с.
- [4] Рекуррентные нейронные сети: типы, обучение, примеры и применение [Электронный ресурс]. – Режим доступа: <https://neurohive.io/ru/osnovy-data-science/rekurrentnye-nejronnye-seti/> (26.12.2022).