

Математическое моделирование процесса оценки стоимости аренды недвижимости с использованием искусственных нейронных сетей

В.А. Евсин, Е.А. Продан, В.А. Евсина, С.Н. Широбокова

*Южно-Российский государственный политехнический университет (НПИ)
имени М.И. Платова, Новочеркасск*

Аннотация: В данной статье рассматривается проблема определения стоимости аренды недвижимости. Обосновывается идея о минимизации функции абсолютной ошибки с использованием искусственных нейронных сетей. Особое внимание уделено процессу определения входных данных нейронной сети. В частности, рассмотрена проблематика определения таких параметров, как благоустройство региона и помещения. В статье выяснены особенности определения весовых коэффициентов для определения технического оснащения помещения с использованием генетического алгоритма. Предложена модель архитектуры нейронной сети. Описана модель изменения весовых коэффициентов. По итогу произведена апробация модели на тестовых данных, а также описана модель корректировки данных с учетом динамики цен.

Ключевые слова: нейронная сеть, *data mining*, анализ данных, аренда недвижимости, регрессия, генетический алгоритм, информатика, машинное обучение, оценка стоимости, моделирование, экстраполяция.

Оценка недвижимости в настоящее время получает все большее значение для поддержания конкурентоспособности на рынке агентств недвижимости. Основной группой методов при проведении оценки являются методы сравнения, в ходе которых производится оценка недвижимости с точки зрения сравнения ее характеристик с существующими оцененными объектами, подробнее о которых в [1]. Для проведения подобной оценки может быть использован аппарат искусственных нейронных сетей [2-4]. Функционал качества прогноза в данном случае может быть основан на минимизации суммы квадратов отклонений и иметь следующую форму:

$$J(w) = \frac{1}{2} \sum_{j=1}^m [C^T - \tilde{C}^T]^2 \rightarrow \min,$$

где C^T – стоимость объекта фактическая, \tilde{C}^T – стоимость объекта прогнозируемая, которая может быть выражена следующей формулой:

$$\tilde{C}^T = IH * w_1 + IR * w_2 + DC * w_3 + M * w_4 + TS * w_5 + LN * w_6 + LT * w_7 + \varepsilon,$$

где IH – коэффициент благоустройства помещения, IR – коэффициент благоустройства региона, DC – удаленность от центра, M – материал помещения, TS – коэффициент технического оснащения квартиры, LN – долгота, LT – широта, ε – случайная ошибка модели. С учетом представленных параметров функционал приобретает следующую форму:

$$J(w) = \frac{1}{2} \sum_{j=1}^m [C^T - (IH * w_1 + IR * w_2 + DC * w_3 + M * w_4 + TS * w_5 + LN * w_6 + LT * w_7)] \rightarrow \min,$$
 параметры которой идентичны представленным ранее.

Для выявления интегральной оценки влияния t -го типа объектов инфраструктуры региона на стоимость объекта недвижимости необходимо вывести дифференциальное уравнение, описывающее данный процесс. Рассматривая связь между приращением аргумента ΔK_i^t , который представлен объектами t -го типа инфраструктуры региона, и приращением влияния ΔIR_i , можно вывести зависимость. Для восстановления данных следует провести процедуру интерполяции данных формулой Ньютона, подробнее о которой в [5,6], при которой зависимость будет иметь следующий вид:

$$IR^t(K) = IR_0^t + \sum_{i=1}^n IR^t(K_0, \dots, K_i) * (K - K_0) * \dots * (K - K_{i-1}),$$

где $IR^t(K)$ – функциональная зависимость влияния от количества объектов данного типа, IR_0^t – значение зависимости при отсутствии объектов данного типа, $IR^t(K_0, \dots, K_i)$ – разделенная разность i -го порядка для характеристики влияния. Для определения коэффициента благоустройства региона от влияния объектов t -го типа используется следующая зависимость:

$$IR = \sum_{t=1}^m M_t * e^{-a * r_t} * Ln(K_t),$$

где M – коэффициент влияния от объектов t -го типа, a – функция насыщения объектами t -го типа, r – усредненное расстояние до объектов, K – количество значимых объектов региона.

Интегральная оценка благоустроенности помещения складывается из оценки характеристических показателей помещения, а также показателей состояния предметов в помещении с учетом их качества и имеет вид:

$$IH = \sum_{s=1}^r W_{1,s} * C_{1,s} + \sum_{i=1}^n W_{2,i} * C_{2,i} * (1 - a_{2,i}),$$

где IH – коэффициент благоустроенности помещения, $W_{1,s}$ – весовой коэффициент s -го характеристического показателя, $C_{1,s}$ – степень качества s -го характеристического показателя, $W_{2,i}$ – весовой коэффициент от наличия i -го предмета, $C_{2,i}$ – степень ценности i -го предмета, $a_{2,i}$ – коэффициент износа. Коэффициент технического оснащения квартиры включает следующие данные: газовое обеспечение, наличие интернета, холодное водоснабжение, горячее водоснабжение, водоотведение, электричество. Для определения весовых коэффициентов от наличия каждого из компонент необходимо определить следующий функционал качества:

$$\left\{ \begin{array}{l} J(w) = \frac{1}{2} \sum_{j=1}^m [C^T - (k_1 * x_1 + k_2 * x_2 + \dots + k_7 * x_7)]^2 \rightarrow \min \\ 0 < x_i < 1 \\ \sum_i x_i = 1 \end{array} \right. ,$$

где $i = \overline{1, n}$ – компоненты технического обеспечения, x_i – весовой коэффициент i -й компоненты, k_i – коэффициент наличия i -й компоненты, причем:

$$k_i = \begin{cases} 1, & \text{если помещение обеспечено } i\text{-м компонентом} \\ 0, & \text{иначе} \end{cases} .$$

Данный функционал качества исходит из предположения, что наличие каждой компоненты положительно сказывается на стоимости недвижимости. Для решения данного функционала может быть использован аппарат генетических алгоритмов [7].

Хромосома может быть представлена в следующей форме:

$$Ch = [b_1; b_2; \dots; b_{48}; b_{49}], \| b_i \| = \{0, 1\} .$$

В качестве начальной популяции принимаются наборы бинарных данных, которые в десятичном представлении формируют допустимое множество решений. Фитнес-функция модели имеет следующую форму:

$$FF = \sum_{i=0}^m (C^T_i - \sum_j \frac{Dec(BinArray_j)}{100} * p_{i,j}),$$

где FF – фитнес-функция модели, $Dec(BinArray_j)$ – десятичное представление j -го параметра хромосомы, $p_{i,j}$ – значение j -го параметра i -й строки набора данных. В качестве модели селекции используется модель элитарного отбора по критерию минимальной фитнес-функции хромосом. Для определения генов новых потомков используется модель четырнадцати точечного кроссовера. Кроме того, с вероятностью 0.7 возникает мутация у рассматриваемой хромосомы. Условие выхода из процесса селекции имеет следующую форму:

$$Ex = \begin{cases} 1, \text{ если } \frac{1}{n} \sum_j |FFB_j - \overline{FFB}| < \varepsilon \text{ или } c_ep = m_ep, \\ 0, \text{ иначе} \end{cases}$$

где FFB_j – наиболее приспособленная особь j -й эпохи, \overline{FFB} – усредненный показатель наиболее приспособленных особей за последние n -эпох, ε – максимальная ошибка модели, c_ep – номер текущей эпохи, m_ep – максимальное количество эпох.

При построении нейронной сети используется оптимизация методом стохастического градиентного спуска, который имеет следующую форму:

$$\begin{aligned} \delta_t &= \nabla_{w_t} J(w_t), \\ m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \delta_t \\ \hat{v}_t &= \beta_2 \hat{v}_{t-1} + (1 - \beta_2) |\delta_t|^2 \\ w_{t+1} &= w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon} (\beta_1 w_t - \frac{\beta_1 \eta}{1 - \beta_1^t} + \frac{(1 - \beta_1^t) \delta_t}{1 - \beta_1^t}) \end{aligned}$$

где δ_t – градиент, m_t – момент расчета, η – темп обучения, β_1, β_2 – параметры обучения.

Сеть имеет входной, выходной слой, а также 10 скрытых слоев с конфигурацией количества нейронов в каждом: 7/7/64/128/64/8/6/5/4/3/2/1.

В качестве функции активации используется модель *SELU* (масштабная экспоненциальная линейная единица), которая имеет следующую форму:

$$SELU(x) = \lambda \begin{cases} x, & \text{если } x > 0 \\ ae^x - a, & \text{иначе} \end{cases}$$

В результате обучения модели сформирован график динамики средней абсолютной ошибки обучения, представленный на рис. 1. Результаты проверки модели на тестовых данных представлены на рис. 2.

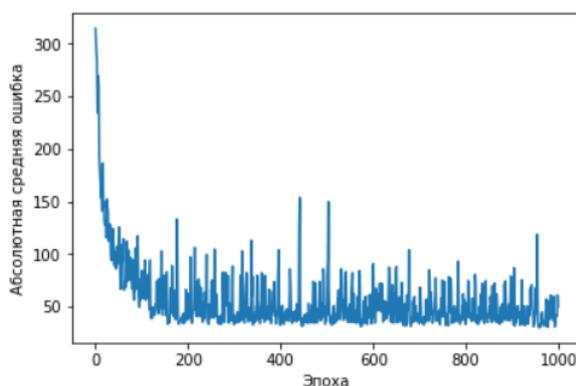


Рис. 1. – Динамика изменения абсолютной средней ошибки обучения

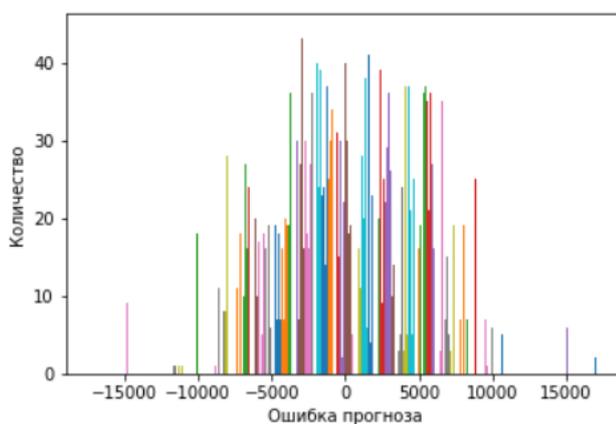


Рис. 2. – Диаграмма распределения ошибки прогноза

После обучения модели необходимо провести корректировку полученных результатов с учетом динамики средних цен на недвижимость по периодам. Для проведения данной корректировки может быть

использована адаптивная модель прогнозирования Брауна второго порядка, подробнее о которой в [8] и которая имеет следующую форму:

$$M(t+k) = A_0 + A_1 * k + A_2 * k^2,$$

где t – текущее время, k – время упреждения. Порядок модели определяется априорно из предварительного анализа временного ряда и законов развития динамики рынка недвижимости, A_0, A_1, A_2 – коэффициенты регрессии. Для формирования модели необходимо определить коэффициенты методом парной регрессии, после чего итеративно исчислять ошибку прогнозирования и корректировать значения коэффициентов. Модель корректировки коэффициентов на каждой итерации имеет следующую форму:

$$\begin{aligned}\tilde{Y}(t+1) &= A_0 + A_1 * (t+1) + A_2 * (t+1)^2 \\ e(t+1) &= \text{abs}(Y(t+1) - \tilde{Y}(t+1)) \\ A_0 &= A_0(t) + A_1 * (t+1) + A_2 * (t+1)^2 + (1-\beta)^2 * e(t+1), \\ A_1 &= A_1(t+1) + A_2 * (t+1)^2 + (1-\beta)^2 * e(t+1) \\ A_2 &= A_2 * (t+1)^2 + (1-\beta)^2 * e(t+1)\end{aligned}$$

где $\tilde{Y}(t+1)$ – прогнозируемое значение в $(t+1)$ периоде, $Y(t+1)$ – реальное значение в $(t+1)$ периоде, e – ошибка модели, β – коэффициент обучения. Данная модель позволит провести уточнение стоимости недвижимости с учетом динамики цен.

Проектируемый модуль может быть использован в информационных системах по управлению недвижимостью, в частности, в распределенном реестре, применение искусственных нейронных сетей и математическое моделирование которого представлено в [9,10]. Дальнейшее проектирование модуля будет направлено на совершенствование модели нейронной сети.

Литература

1. Симионова Н. Е., Шеина С.Г. Методы оценки и технической экспертизы недвижимости. М.: ИКЦ "МарТ", 2006. 448 с.



2. Хайкин С. Нейронные сети. Полный курс. 2 изд. М.: Издательский дом "Вильямс", 2006. 1104 с.
 3. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. М.: Горячая линия. Телеком, 2002. 382 с.
 4. Dean J. Big Data, Data Mining, and Machine Learning. Value Creation for Business Leaders and Practitioners. New Jersey: John Wiley & Sons Limited, 2014. 208 p.
 5. Gautschi W. Numerical Analysis. 2 edition. Basel: Birkhäuser, 2012. 615 p.
 6. Богуславский И.А. Полиномиальная аппроксимация для нелинейных задач оценивания и управления. 2 изд. М.: Физматлит, 2006. 208 с.
 7. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. М.: Горячая линия – Телеком, 2006. 383 с.
 8. Munbaev K. Short-Memory Linear Processes and Econometric Applications. New Jersey: John Wiley & Sons Limited, 2011. 451 p.
 9. Евсин В.А., Широбокова С.Н., Продан Е.А. Использование технологии распределенных реестров при проектировании информационной системы «Аренда недвижимости» с применением искусственных нейронных сетей // Инженерный вестник Дона, 2018, №1. URL: ivdon.ru/ru/magazine/archive/n1y2018/4655.
 10. Евсин В.А., Широбокова С.Н., Евсина В.А., Продан Е.А. Математическое моделирование распределенного реестра в сфере аренды недвижимости как сети массового обслуживания // Инженерный вестник Дона, 2018, №3. URL: ivdon.ru/ru/magazine/archive/n3y2018/5078.
-

References

1. Simionova N. E., Sheina S.G. Metody otsenki i tekhnicheskoy ekspertizy nedvizhimosti [Methods of evaluation and technical expertise of real estate]. M.: IKTs "MarT", 2006. 448p.
2. Hajkin S. Nejronnye seti. Polnyj kurs. [Neural networks. A Comprehensive Foundation]. 2 izd. M.: Izdatel'skij dom "Vil'jams", 2006. 1104p.
3. Kruglov V.V., Borisov V.V. Iskusstvennye nejronnye seti. Teorija i praktika. [Artificial neural network. Theory and practice.] M.: Gorjachaja linija. Telekom, 2002. 382p.
4. Dean J. Big Data, Data Mining, and Machine Learning. Value Creation for Business Leaders and Practitioners. New Jersey: John Wiley & Sons Limited, 2014. 208p.
5. Gautschi W. Numerical Analysis. 2 edition. Basel: Birkhäuser, 2012. 615p.
6. Boguslavskiy I.A. Polinomial'naya approksimatsiya dlya nelineynykh zadach otsenivaniya i upravleniya [Polynomial approximation for nonlinear estimation and control problems]. 2 izd. M.: Fizmatlit, 2006. 208 p.
7. Rutkovskaja D., Pilin'skij M., Rutkovskij L. Nejronnye seti, geneticheskie algoritmy i nechetkie sistemy. [Neural networks, genetic algorithms and fuzzy systems] M.: Gorjachaja linija. Telekom, 2006. 383 p.
8. Munbaev K. Short-Memory Linear Processes and Econometric Applications. New Jersey: John Wiley & Sons Limited, 2011. 451 p.
9. Evsin V.A., Shirobokova S.N., Prodan E.A. Inženernyj vestnik Dona (Rus), 2018, №1. URL: ivdon.ru/ru/magazine/archive/n1y2018/4655.
10. Evsin V.A., Shirobokova S.N., Evsina V.A., Prodan E.A. Inženernyj vestnik Dona (Rus), 2018, №3. URL: ivdon.ru/ru/magazine/archive/n3y2018/5078.