DOI: 10.12731/2070-7568-2021-10-4-171-180
УДК 004.415.2

ПРИМЕНЕНИЕ МАШИННЫХ АЛГОРИТМОВ ДЛЯ ПРОГНОЗИРОВАНИЯ СТОИМОСТИ НЕДВИЖИМОСТИ

Павлова А.И., Корж А.А.

Цель *– анализ методов машинного обучения для прогнозирования стоимости жилой недвижимости.*

Метод или методология проведения работы: *в статье использованы методы машинного обучения обучения глубоких нейронных сетей: стохастический градиентный спуск (SGD), метод адаптивного градиента (Adagrad), метод адаптивного скользящего среднего градиентов (RMSprop), метод адаптивного шага обучения (Adadelta), метод Адама (Adam).*

Результаты: *построена модель обучения нейронной сети для прогнозирования стоимости жилой недвижимости. В качестве предикторов использована информация о площади земельного участка, количестве спален, количество и качество ванных комнат, оценку общего качества жилья, оценку состояния жилой недвижимости, количество каминов, площади гаража, общее количество комнат. Анализ точности алгоритмов машинного обучения показал, что меньшие ошибки получены при использовании метода адаптивного скользящего среднего градиентов (RMSprop).*

Область применения результатов: *полученные результаты целесообразно применять при прогнозировании стоимости жилой недвижимости.*

Ключевые слова: *искусственные нейронные сети; градиент; недвижимость; прогнозирование*

APPLICATION OF MACHINE ALGORITHMS FOR FORECASTING REAL ESTATE COSTS

Pavlova A.I., Korzh A.A.

Purpose *– development of web-application using the system of automated interaction with enterprise clients (Customer Relationship Management, CRM-system), aimed at conducting anti-collector activity.*

Method or methodology of the work: *programming methods were used in the article.*

Results: *the web application for the management of the anti-collectors' activity integrated with the CRM system Bitrix 24 was developed.*

The sphere of application of the results: *the received results are to be applied to the management of the activity of anti-collector enterprises which is connected with structuring and liquidation of debts of physical and legal persons.*

Keywords: *anti-collector activity; web-application; information system; enterprise management system*

Введение

Искусственные нейронные сети находят применение при решении сложных задач, когда обычные алгоритмические решения оказываются неэффективными или невозможными. При построении нейронных сетей делается ряд допущений и значительных упрощений, однако, они демонстрируют такие свойства, как обучение на основе опыта, обобщение, извлечение существенных данных из избыточной информации. После анализа входных сигналов нейронные сети способны к самообучению. Способность к моделированию нелинейных процессов, работе с зашумленными данными и адаптивность дают возможности применять нейронные сети для решения широкого класса финансовых задач [1-3]. Одной из актуальных проблем является прогнозирование стоимости жилой недвижимости. Применяемые для прогнозирования методы разнообразны. По мнению авторов [4-5] наибольшей прогностической способностью обладают методы, включающие в себя подходы эвристического и статистического анализа данных.

Целью работы является анализ методов машинного обучения для прогнозирования стоимости жилой недвижимости.

Материалы и методы работы

В качестве исходных данных были использованы сведения о стоимости недвижимости, состоящий из 1460 строк и 10 столбцов [6]. Набор данных содержит информацию о площади земельного участ-

ка, количестве спален, ванных комнат, об оценке качества жилья и состояния жилой недвижимости, о количестве каминов, площади гаража, количестве комнат.

В зарубежной литературе градиентные алгоритмы широко используются для построения модели прогнозирования стоимости недвижимости [7-11].

При построении модели прогнозирования стоимости жилой недвижимости использованы известные алгоритмы обучения глубоких нейронных сетей: стохастический градиентный спуск (SGD) [12-13], метод адаптивного градиента (Adagrad) [14], метод адаптивного скользящего среднего градиентов (RMSprop) [15], метод адаптивного шага обучения (Adadelta) [16], метод Адама (Adam) [17].

Результаты исследований

С использованием библиотек Sklearn [18] машинного обучения и платформы TensorFlow [19] с открытым исходным кодом для машинного обучения создана модель прогнозирования стоимости жилой недвижимости с входными признаками, приведенными на рис. 1. В качестве выходного прогнозируемого значения в исходном наборе данных служил показатель AboveMedinaPrice, представленный в двоичном виде (значение 1 соответствует оценке стоимости недвижимости вышей средней рыночной, а значение 0 соответствует оценке стоимости недвижимости ниже средней рыночной).

	LotArea	OverallQual	OverallCond	TotalBsmtSF	FullBath	HalfBath	BedroomAbvGr	TotRmsAbvGrd	Fireplaces	GarageArea	AboveMedianPrice
0	8450	7	5	856	2	1	3	8	0	548	1
1	9600	6	8	1262	2	0	3	6	1	460	1
2	11250	7	5	920	2	1	3	6	1	608	1
3	9550	7	5	756	1	0	3	7	1	642	0
4	14260	8	5	1145	2	1	4	9	1	836	1

Рис. 1. Входные признаки для построения модели прогнозирования

На рис. 2 приведены статистические показатели входных признаков (count – количество примеров, mean – среднее значение,

std – стандартное отклонение, min – минимальное, 25%, 50% и 75% процентили, max –максимальное значение). Общее количество примеров в наборе данных составило 1460.

	LotArea	OverallQual	OverallCond	TotalBsmtSF	FullBath	HalfBath	BedroomAbvGr	TotRmsAbvGrd	Fireplaces	GarageA
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000
mean	10516.828082	6.099315	5.575342	1057.429452	1.565068	0.382877	2.866438	6.517808	0.613014	472.980
std	9981.264932	1.382997	1.112799	438.705324	0.550916	0.502885	0.815778	1.625393	0.644666	213.804
min	1300.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	2.000000	0.000000	0.0000
25%	7553.500000	5.000000	5.000000	795.750000	1.000000	0.000000	2.000000	5.000000	0.000000	334.500
50%	9478.500000	6.000000	5.000000	991.500000	2.000000	0.000000	3.000000	6.000000	1.000000	480.000
75%	11601.500000	7.000000	6.000000	1298.250000	2.000000	1.000000	3.000000	7.000000	1.000000	576.000
max	215245.000000	10.000000	9.000000	6110.000000	3.000000	2.000000	8.000000	14.000000	3.000000	1418.000

Рис. 2. Статистические показатели входных признаков

Модель реализована в виде глубокой нейронной сети с двумя скрытыми слоями. В качестве функций активации нейронов использованы сигмоидальная (Sigmoid) для выходного слоя и ReLu для промежуточных слоев. Обучение модели прогнозирования осуществлюсь на множестве данных, разделенных на три группы: обучающее (70% от общего числа примеров), тестовое (15%) и вариационное (15%).

Обучение модели прогнозирования выполнено с применением градиентных методов SGD, Adagrad, RMSprop, Adadelta, Adam. Скорость обучения нейронной сети была задана 0.001, количество эпох обучения принято равным 100.

OverallQual	OverallCond	TotalBsmtSF	FullBath	HalfBath	BedroomAbvGr	TotRmsAbvGrd	Fireplaces	GarageArea	AboveMedianPrice	Predict
7	5	856	2	1	3	8	0	548	1	1.106342
6	8	1262	2	0	3	6	1	460	1	1.512457
7	5	920	2	1	3	6	1	608	1	1.453659
7	5	756	1	0	3	7	1	642	0	0.000000
8	5	1145	2	1	4	9	1	836	1	1.743343

Рис. 3. Результаты обучения модели прогнозирования

Для оценки точности использованы показатели: общая оценка обучения, максимальная ошибка, средняя абсолютная ошибка, средняя квадратичная ошибка и медианная абсолютная ошибка (таблица 1).

Таблица 1.

Результаты оценки точности нейронной сети

Результат обучения	Метод обучения сети				
	SGD	Adagrad	RMSprop	Adadelta	Adam
Общая оценка обучения	63%	85%	96%	90%	83%
Максимальная ошибка	0,65	1,97	1,94	1,65	1,97
Средняя абсолютная ошибка	0,47	0,79	0,66	1,00	0,65
Средняя квадратичная ошибка	0,22	0,72	0,66	1,25	0,65
Медианная абсолютная ошибка	0,47	0,97	0,94	0,62	0,97

Результаты оценки точности показали, что метод RMSprop имеет лучшую сходимость, общая оценка обучения составила 96%. При использовании методы SGD общая оценка обучения составила 63%. По другим показателям метод SGD характеризуется меньшими значениями ошибок в сравнении с другими Adadelta, Adagrad, Adam. В целом метод RMSprop характеризуется лучшей сходимостью. Значения средней квадратической ошибки меньше, чем у методов Adadelta, Adagrad, Adam.

В таблице 2 приведены достоинства и недостатки методов оптимизации, использованные в настоящей работе.

Таблица 2.

Общие характеристики методов оптимизации

Метод	Достоинства	Недостатки
SGD	Метод приспособлен для динамического обучения. Алгоритм способен обучаться на избыточно больших выборках.	Алгоритм может не сходиться или сходиться слишком медленно. При большой размерности пространства признаков возможно переобучение, обучение сети может происходить нестабильно.
Adagrad	Регулируется скорость обучения, используется кумулятивная сумма квадратов градиента. Это улучшает производительность при проблемах с разреженными градиентами	Скорость обучения может уменьшаться с течением времени до бесконечно малой величины.
Adadelta	Усовершенствованная версия Adagrad, характеризуется лучшей сходимостью в сравнении с Adagrad и SGD	Возможно переобучение модели или паралич сети.

Окончание табл. 2.

RMSprop	Характеризуется хорошей сходимостью. Алгоритм хорошо работает с онлайн-обучением больших данных.	Скорость обучения адаптируется на основе среднего первого момента (среднего значения). RMSprop вносит свой вклад в экспоненциально затухающее среднее значение прошлых «квадратичных градиентов». Средняя медианная ошибка, средняя абсолютная ошибка и средняя квадратичная ошибка примерно равны с ошибками для метода Adam, но больше чем у стохастического метода SGD.
Adam	Метод, использующий адаптивную скорость обучения, сочетающий подходы Adadelta и RMSprop. Метод имеет хорошую сходимость в сравнении со стохастической оптимизацией, использует среднее значение вторых моментов градиентов. В частности, алгоритм вычисляет экспоненциальное скользящее среднее градиента и квадратичный градиент	Общая оценка метода для тестовой, обучающей и вариационной выборок получена меньше в сравнении методами RMSprop, Adadelta, Adagrad.

Заключение

Методы машинного обучения позволяют строить модели прогнозирования стоимости жилой недвижимости. При этом известные методы обучения, основанные на поиске градиента функции потерь Adadelta, Adagrad, Adam, SGD и RMSprop обладают разной прогностической способностью. Сравнительный анализ результатов обучения данными методами, показал, что наибольшие ошибки в обучении модели возникли при использовании метода SGD (общая оценка обучения составила 63%). Это объясняется случайным поиском направления градиента функции. Однако общие показатели точности метода SGD (максимальная ошибка, средняя абсолютная ошибка, средняя квадратичная ошибка и медианная абсолютная ошибка) наименьшие.

Метод обучения RMSprop имеет высокую прогностическую способность (96%), а значения средней квадратической ошибки меньше, чем у методов Adadelta, Adagrad, Adam.

Список литературы

1. Осовский С. Нейронные сети для обработки информации, пер. с польск. И.Д. Рудинского, 2002. С. 345.
2. Хайкин С. Нейронные сети. М.: Издательский дом Вильямс, 2006. 1001 с.
3. Ежов А. А. Нейрокомпьютинг и его применение в экономике и бизнесе: учеб. пособие / А. А. Ежов, С. А. Шумский. М.: МИФИ, 1998. 224 с.
4. Стерник С.Г., Стерник Г.М. Методика прогнозирования объемов ввода на локальном рынке строительства и продажи жилья // Жилищные стратегии. 2018. Т.5. N 2. С.138-152.
5. Стерник Г.М., Стерник С.Г., Свиридов А.В. Методология прогнозирования российского рынка недвижимости. Ч. 3. Эволюция методов прогнозирования на рынке жилой недвижимости России // Механизация строительства, 2014. № 2(836). С. 60-64.
6. Housepise. https://www.kaggle.com/moewie94/housepricedata?select=housepricedata.csv
7. Caplin A., Chopra S., Leahy J., LeCun Y., Thampy T. Machine Learning and the Spatial Structure of House Prices, 2016, 163 p.
8. Winky K.O. Ho, Bo-Sin Tang, Siu Wai Wong Predicting property prices with machine learning algorithms // Journal of Property Research. 2021. Vol. 38. No. 1. P. 48-70. DOI: https://doi.org/10.1080/09599916.2020.1832558
9. Gu, J., Zhu, M., & Jiang, L. Housing price forecasting based on genetic algorithm and support vector machine // Expert System with Applications. 2011. Vol. 38(4). P. 3383–3386. DOI: https://doi.org/10.1016/j.eswa.2010.08.123
10. Limsombunchai, V., Gan, C., Lee M. House price prediction: Hedonic price model vs. artificial neural network // American Journal of Applied Sciences. 2000. Vol. 1(3). P. 193–201. DOI: https://doi.org/10.3844/ajassp.2004.193.201
11. Mu, J. Y., Wu, F., Zhang A. H. Housing value forecasting based on machine learning methods // Abstract and Applied Analysis. 2014. Article ID 648047. DOI: http://dx.doi.org/10.1155/2014/648047

12. Созыкин А.В. Обзор методов обучения глубоких нейронных сетей // Вестник ЮУрГУ. Сер. Вычислительная математика и информатика. 2017. Т. 6, № 3. С. 28-59.
13. Robinds H., Monro S. A stochastic approximation method // Annals of Mathematical Statistics. 1951. Vol. 22. P. 400-407.
14. Zhang Y. R., Haghani, A. A gradient boosting method to improve travel time prediction // Transportation Research Part C: Emerging Technologies. 2015. No. 58. P. 308–324. DOI: https://doi.org/10.1016/j.trc.2015.02.019
15. Christian Igel and Michael Hüsken Improving the Rprop Learning Algorithm. 2000. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.17.1332
16. Zeiler M. D. ADADELTA: An Adaptive Learning Rate Method // Retrieved, 2021. http://arxiv.org/abs/1212.5701
17. Kingma D. P., Ba J. L. Adam: a Method for Stochastic Optimization // International Conference on Learning Representations. 2017. P. 1-13. https://arxiv.org/pdf/1412.6980.pdf
18. An introduction to machine learning with scikit-learn. https://scikit-learn.org/stable/tutorial/basic/tutorial.html#machine-learning-the-problem-setting
19. Tensorflow. https://www.tensorflow.org/

References

1. Osovskiy S. *Neyronnye seti dlya obrabotki informatsii* [Neural networks for information processing]. 2002, p. 345.
2. Khaykin S. *Neyronnye seti* [Neural networks]. M.: Izdatel'skiy dom Vil'yams, 2006, 1001 p.
3. Ezhov A. A. Neyrokomp'yuting i ego primenenie v ekonomike i biznese: ucheb. posobie / A. A. Ezhov, S. A. Shumskiy. M.: MIFI, 1998. 224 s.
4. Sternik S.G., Sternik G.M. *Zhilishchnye strategii,* 2018, vol. 5, no. 2, pp. 138-152.
5. Sternik G.M., Sternik S.G., Sviridov A.V. *Mekhanizatsiya stroitel'stva*, 2014, no. № 2(836), pp. 60-64.

6. Housepise. https://www.kaggle.com/moewie94/housepricedata?select=-housepricedata.csv
7. Caplin A., Chopra S., Leahy J., LeCun Y., Thampy T. Machine Learning and the Spatial Structure of House Prices, 2016, 163 p.
8. Winky K.O. Ho, Bo-Sin Tang, Siu Wai Wong Predicting property prices with machine learning algorithms. *Journal of Property Research,* 2021, vol. 38, no. 1, pp. 48-70. DOI: https://doi.org/10.1080/09599916.2020.1832558
9. Gu, J., Zhu, M., & Jiang, L. Housing price forecasting based on genetic algorithm and support vector machine. *Expert System with Applications,* 2011, vol. 38(4), pp. 3383–3386. DOI: https://doi.org/10.1016/j.eswa.2010.08.123
10. Limsombunchai, V., Gan, C., Lee M. House price prediction: Hedonic price model vs. artificial neural network. *American Journal of Applied Sciences*, 2000, vol. 1(3), pp. 193–201. DOI: https://doi.org/10.3844/ajassp.2004.193.201
11. Mu, J. Y., Wu, F., Zhang A. H. Housing value forecasting based on machine learning methods. *Abstract and Applied Analysis,* 2014. Article ID 648047. DOI: http://dx.doi.org/10.1155/2014/648047
12. Sozykin A.V. *Vestnik YuUrGU. Ser. Vychislitel'naya matematika i informatika*, 2017, vol. 6, no. 3, pp. 28-59.
13. Robinds H., Monro S. A stochastic approximation method. *Annals of Mathematical Statistics,* 1951, vol. 22, pp. 400-407.
14. Zhang Y. R., Haghani, A. A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 2015, no. 58, pp. 308–324. DOI: https://doi.org/10.1016/j.trc.2015.02.019
15. Christian Igel and Michael Hüsken Improving the Rprop Learning Algorithm. 2000. http://citeseerx.ist.psu.edu/viewdoc/summary?-doi=10.1.1.17.1332
16. Zeiler M. D. ADADELTA: An Adaptive Learning Rate Method. Retrieved, 2021. http://arxiv.org/abs/1212.5701
17. Kingma D. P., Ba J. L. Adam: a Method for Stochastic Optimization. *International Conference on Learning Representations,* 2017, pp. 1-13. https://arxiv.org/pdf/1412.6980.pdf

18. An introduction to machine learning with scikit-learn. https://scikit-learn.org/stable/tutorial/basic/tutorial.html#machine-learning-the-problem-setting
19. Tensorflow. https://www.tensorflow.org/

ДАННЫЕ ОБ АВТОРАХ

Павлова Анна Илларионовна, кандидат технических наук, доцент
Новосибирский государственный университет экономики и управления
ул. Каменская, 565, г. Новосибирск, 630039, Российская Федерация
annstab@mail.ru

Корж Александр Александрович, студент
Новосибирский государственный университет экономики и управления
ул. Каменская, 565, г. Новосибирск, 630039, Российская Федерация

DATA ABOUT THE AUTHORS

Anna I. Pavlova, PhD (technical sciences), associate professor
Novosibirsk State University of Economics and Management
56, Kamenskaya Str., Novosibirsk, 630039, Russian Federation
annstab@mail.ru
SPIN-code: 8714-1140
ORCID: 0000-0001-6159-1439
Scopus Author ID: 0000-0001-6159-1439

Alexander A. Korzh, student
Novosibirsk State University of Economics and Management
56, Kamenskaya Str., Novosibirsk, 630039, Russian Federation

Поступила 01.11.2021 Received 01.11.2021
После рецензирования 16.11.2021 Revised 16.11.2021
Принята 22.11.2021 Accepted 22.11.2021