

Оптимизация нейросетевого алгоритма описания земельного рынка

Мирон А. Карпович,¹

Екатерина Г. Фролова¹ eg.frolova@gkp.vrn.ru

¹кафедра экономики и основ предпринимательства, Воронеж. гос. архитектурно-строительный ун-т, ул. 20 летия Октября, 64, г. Воронеж, 394006, Россия

Реферат. Показаны преимущества использования нейросетевых технологий по сравнению с традиционными при описании динамически меняющихся систем, к которым относится и современный земельный рынок. Выявлена основная трудность, возникающая во время практических реализаций нейросетевых моделей рынка земли и строительной продукции, состоящая в формировании репрезентативного набора обучающих и тестовых примеров. Определены требования, выполнение которых необходимо для корректного описания современной экономической ситуации, заключающееся в том, что тренировочное множество в пространстве признаков не должно иметь областей с малой плотностью наблюдений. Сформулированы методы оптимизации эмпирического массива, позволяющие избежать далёкой экстраполяции данных из областей сгущения набора примеров. Показано, что радикальный метод оптимизации набора обучающих и тестовых примеров, заключающийся в сборе дополнительной информации, для экономических задач обычно сопряжён со значительными затратами ресурсов и времени и по соотношению затраты/отдача менее эффективен, чем алгоритм оптимизации нейросетевых моделей земельного рынка на фиксированном наборе эмпирической информации. Подробно проанализирован алгоритм оптимизации, основанный на преобразовании массивов информации и заключающийся в растяжении областей сгущения набора примеров, а также сжатия областей малой плотности наблюдений. На конкретном примере рынка земельных участков Воронежской области, предназначенных для дорожного строительства, продемонстрировано значительное уменьшение относительной ошибки описания цены земельного участка, что с учётом значительной абсолютной стоимости земли делает рентабельным применение даже радикального метода оптимизации эмпирического массива. Продemonстрирована высокая экономическая эффективность применения предложенных алгоритмов.

Ключевые слова: рынок земли, дорожное строительство, нейронная сеть, оптимизация, экономическая эффективность

Optimization of neural network algorithm of the land market description

Myron A. Karpovich,¹

Ekaterina G. Frolova¹ eg.frolova@gkp.vrn.ru

¹economics and the basics of entrepreneurship department, Voronezh. state. Architecture and Construction University, 20 Ocityabya str, 64, Voronezh, 394006, Russia

Summary. The advantages of neural network technology is shown in comparison of traditional descriptions of dynamically changing systems, which include a modern land market. The basic difficulty arising in the practical implementation of neural network models of the land market and construction products is revealed. It is the formation of a representative set of training and test examples. The requirements which are necessary for the correct description of the current economic situation has been determined, it consists in the fact that Train-paid-set in the feature space should not have the ranges with a low density of observations. The methods of optimization of empirical array, which allow to avoid the long-range extrapolation of data from range of concentration of the set of examples are formulated. It is shown that a radical method of optimization a set of training and test examples enclosing to collect supplementary information, is associated with significant costs time and resources for the economic problems and the ratio of cost / efficiency is less efficient than an algorithm optimization neural network models the earth market fixed set of empirical data. Algorithm of optimization based on the transformation of arrays of information which represents the expansion of the ranges of concentration of the set of examples and compression the ranges of low density of observations is analyzed in details. The significant reduction in the relative error of land price description is demonstrated on the specific example of Voronezh region market of lands which intend for road construction, it makes the using of radical method of empirical optimization of the array cost-effective with accounting the significant absolute value of the land. The high economic efficiency of the proposed algorithms is demonstrated.

Keywords: land market, road construction, neural network, optimization, economic efficiency

Введение

На протяжении многих лет в качестве основного метода описания динамики и оптимизации сложных систем использовались линейные модели, для которых разработаны эффективные алгоритмы оптимизации. Однако для большинства реальных систем (и, в частности, экономических) линейная аппроксимация неадекватна реальности [1].

Для цитирования

Карпович М. А., Фролова Е. Г. Оптимизация нейросетевого алгоритма описания земельного рынка // Вестник ВГУИТ. 2016. № 2. С. 293–298. doi:10.20914/2310-1202-2016-2-293-298

В связи с этим в последнее время наблюдается рост интереса к нейронным сетям: они успешно используются в самых различных областях науки и практики, в том числе и в экономике. Сферой их применения являются задачи прогнозирования, классификации и управления, для которых отсутствуют адекватные количественные модели или они чересчур сложны для практики.

For citation

Karpovich M. A., Frolova E. G. Optimization of neural network algorithm of the land market description. *Vestnik VSUET* [Proceedings of VSUET]. 2016. no. 2 pp. 293–298. (in Russ.). doi:10.20914/2310-1202-2016-2-293-298

Преимущества использования нейросетевых технологий по сравнению с традиционными определяются, в основном, тремя обстоятельствами. Во-первых, нейронные сети принципиально нелинейны, вследствие чего они представляют собой исключительно мощный метод моделирования, позволяющий воспроизводить сложные зависимости. Во-вторых, нейронные сети эффективно справляются с «проклятием размерности», которое не позволяет моделировать даже и линейные зависимости в случае большого числа переменных [2]. И, в-третьих, они пригодны для описания динамически меняющихся систем, поскольку при неизменной архитектуре адаптируются к быстро меняющейся реальности за счёт изменения только тренировочного набора примеров [3].

Нейросетевые технологии характеризуются возможностью нелинейного моделирования сложных многомерных систем. При этом решение большей части задач, связанных с проектированием, обучением и эксплуатацией нейросетей хорошо известно [4]. Создан разнообразный инструментарий, включающий набор алгоритмов и их программных реализаций, позволяющих даже неспециалисту в области математики и теории управления ставить и решать задачи нейросетевого описания сложных динамических систем [5]. Это позволило, в частности, успешно описывать динамику социально-экономических систем и бизнес-структур (см., например, обзор [6] и цитированную в нём литературу). При этом в сфере экономики наибольшее применение нашли нейросетевые технологии в банковском [7] и биржевом секторах [8]. Для таких систем характерны постоянно динамически обновляющиеся значительные объёмы информации.

В противоположность этому примеры успешного использования нейросетевых алгоритмов к описанию рынков земли и строительной продукции весьма немногочисленны [9], что обусловлено неполнотой и кластеризацией эмпирических данных для таких систем. Главной трудностью для практических реализаций нейросетевых моделей является адекватное реальной ситуации формирование репрезентативного набора обучающих и тестовых примеров [10]. При этом для корректного описания реальной экономической ситуации тренировочное множество в пространстве признаков не должно иметь областей, в котором плотность наблюдений мала. В ином

случае в этих областях будет необходима экстраполяция данных из областей сгущения набора примеров. Такая экстраполяция для нелинейных систем, к которым относятся и социально-экономические, недостоверна [11], вследствие чего могут возникнуть ошибки классификации и описания реальных систем.

Радикальный метод улучшения степени репрезентативности набора обучающих и тестовых примеров, заключающийся в сборе дополнительной информации, более подробно описывающей систему в областях малой плотности наблюдений, для экономических задач обычно сопряжён со значительными затратами ресурсов и времени. Поэтому, как правило, более эффективными по параметру затраты/отдача являются алгоритмы оптимизации количественных моделей систем на фиксированном наборе эмпирической информации. Одним из таких алгоритмов является метод преобразования массивов информации, заключающийся в растяжении областей сгущения набора примеров и сжатия областей малой плотности наблюдений.

Оптимизация нейросетевого описания рынка земельных участков для дорожного строительства

Как свидетельствует анализ данных земельного рынка Воронежской области, исходное распределение наблюдений характеризуется большой неравномерностью: более 34% сосредоточено в пятипроцентной области характеристик с центром в точке 0,71 и ещё более 20% локализованы в десятипроцентной области характеристик с центром в точке 0,35. Суммарно более половины эмпирического массива локализовано в области шириной менее 15% описываемого интервала характеристики. При использовании для обучения нейросети такого массива адекватно будет описана лишь малая часть исследуемых объектов.

Значительно лучше будет обучена нейросеть на массиве преобразованных данных. Как свидетельствуют результаты расчётов, даже максимальный разброс плотностей наблюдений менее двукратного. При этом минимальная плотность преобразованных наблюдений реализуется на границах области описания. На практике объекты с экстремальными характеристиками реализуются реже, чем со средними.

Расчёты свидетельствуют о том, что добавление результатов шести наблюдений в областях их малой плотности значительно улучшает обучаемость сети во всём диапазоне описываемого параметра.

Выбор метода увеличения репрезентативности набора обучающих и тестовых примеров и дальнейшее улучшение обучаемости нейросети в любом случае на первом этапе требует анализа частотного распределения имеющейся информации.

Выполним этот анализ на конкретном примере земельных участков, постоянное и временное занятие которых необходимо для строительства объекта «Реконструкция автомобильной дороги М-29 «Кавказ» – из Краснодара (от Павловской) через Грозный, Махачкалу до границы с Азербайджанской Республикой (на Баку) на участке км 368+000–387+000, Ставропольский край». Всего этот массив информации содержит 149 элементов. Земельные участки описываются следующим набором характеристик:

1. Кадастровый номер выделяемой части.
2. Категория земель.
3. Собственник земельного участка.
4. Наличие инженерных коммуникаций на участке.
5. Плодородие почвы.
6. Удалённость от лесного участка.
7. Удалённость от водного объекта.
8. Удалённость от существующей дороги.
9. Удалённость от населённого пункта.
10. Плодородие почвы.
11. Площадь постоянного отвода.
12. Рыночная стоимость по требованию.
13. Площадь временного отвода.
14. Убытки, упущенная выгода.
15. Стоимость единицы площади.
16. Упущенная выгода с единицы площади.

В наборе характеристик курсивом выделены номинальные переменные, принимающие значения из фиксированного множества. Плотным шрифтом выделены выходные параметры нейросетевой модели.

Среди номинальных переменных характеристика «Категория земель» реально представлены только для земель сельскохозяйственного назначения (свыше 93% примеров), поэтому только такие участки могут быть описаны на этом массиве эмпирической информации. По номинальной характеристике «Собственник земельного участка» реально представлены участки, находящиеся в общей долевой собственности и принадлежащие физическим лицам. Информация об участках, находящихся в госсобственности и собственности юридических лиц в данной выборке недостаточна для их количественной оценки действительных переменных.

Действительные характеристики «Рыночная стоимость по требованию» и «Стоимость единицы площади» не являются независимыми, так же, как и пара характеристик «Убытки, упущенная выгода» и «Упущенная выгода с единицы площади». При этом удельные стоимости и убытки на этом массиве информации по свойствам близки к номинальным переменным: данные для них чётко распадаются на два класса, ширины которых много меньше расстояния между ними. Поэтому нейросетевая задача в этом случае является задачей классификации, а не аппроксимации. Действительная переменная «Плодородие почвы» принимает единственное значение $P=0,93$, вследствие чего только для таких участков с/х назначения может быть выполнена тренировка нейросети. Оценим теперь распределение остальных действительных переменных в массиве эмпирической информации.

Анализ эмпирической информации позволяет сделать следующие выводы:

— Для тех характеристик, у которых значительная часть наблюдений сосредоточена в малой области, эффективным методом оптимизации информации является растяжение области сгущения примеров. Альтернативный путь добавления информации требует весьма значительных затрат;

— Хотя формально и для площадей временного и постоянного отвода также существуют области сгущения при малых значениях соответствующих характеристик, однако, метод растяжения этих областей неадекватен реальной экономической ситуации, что определяется доминирующим вкладом малого числа крупных участков в земельный рынок. Для этих характеристик необходимо подробное исследование сегмента крупных участков, а при невозможности получения такой информации с приемлемыми затратами – кластеризация данных и отдельное описание различных сегментов рынка;

— Для характеристик с достаточно равномерным распределением информации даже небольшое увеличение числа примеров весьма значительно улучшает обучаемость нейросети. Поэтому для таких характеристик эффективен радикальный метод оптимизации информации и лишь при его неприемлемой затратности может быть использован расчётный метод растяжения/сжатия областей характеристик.

Проанализируем изменение степени обучаемости нейросети при растяжении области сгущения примеров по параметрам «Удалённость от лесного участка» и «Удалённость от водного

объекта» при разбиении полного эмпирического массива из 149 элементов на подмассивы: 130 тренировочных примеров и 19 – тестовых.

Обучение нейросети проводилось на модифицированном путём растяжения области сгущения примеров по параметрам «Удалённость от лесного участка» и «Удалённость от водного объекта» массиве наблюдений.

Сравнение результатов обучения нейросети свидетельствует о том, что оптимизация обучающего множества примеров снижает верхнюю границу ошибки с 0,3% до 0,23%. Кроме того, происходит перераспределение погрешностей описания цены внутри диапазона ошибок: снижается доля погрешностей, превышающих 0,2%, за счёт чего возрастает доля погрешностей, лежащих в окрестности точки 0,1%. Дальнейшее снижение верхней границы погрешности описания цены невозможно осуществить без радикального улучшения обучающего массива.

Обучение нейросети проводилось модифицированным методом путём растяжения области сгущения примеров по параметрам «Удалённость от лесного участка» и «Удалённость от водного объекта» массива наблюдений при дополнении тренировочного массива тремя наблюдениями при больших значениях указанных параметров.

Расчёты свидетельствуют, что совместное применение алгоритма растяжения области параметров и добавления в обучающий массив небольшого числа наблюдений, локализованных в области слабого описания исследуемой

системы, удаётся добиться радикального улучшения работы нейросети: верхняя граница погрешности предсказания цены снижается более, чем в два раза и распределение погрешностей стягивается в область малых (не превышающих величины 0,1%) значений. Столь значительное уменьшение относительной ошибки описания цены земельного участка, с учётом значительной абсолютной стоимости земли, может сделать рентабельным процесс сбора дополнительной информации в области недостаточной плотности эмпирических данных.

Экономические результаты уменьшения ошибки оценки земельной составляющей при применении разработанных в данной диссертации алгоритмов приведены в таблице 4. При расчётах предполагалось, что параметры инфляционных процессов за полный жизненный цикл проекта идентичны параметрам за 1995–2015 гг.

Таким образом, дисконтированная выгода Pr , % от уточнения только земельной составляющей бизнес-плана проекта, может быть оценена как:

$$Pr = \frac{14,1 \times (7,3 + 2,1 + 0,7)}{100} = 1,28$$

При цене автодороги сметной стоимостью 100 млн руб. выражение планируемой выгоды составляет 1,28 млн руб. при минимальных затратах на сбор дополнительной информации, выполнение анализа и проведение расчётов.

Таблица 1

Эффективность применения нейросетевых технологий при проведении землеотвода для строительства автомобильных дорог

Table 1

The effectiveness of the use of neural network technology in land allocation for construction of highways

№	Показатель Factor	Оценка показателя% Factor rating, %
1	Доля земельной составляющей The proportion of the land component	9,2
2	Дисконтированная доля земельной составляющей The discounted proportion of the land component	14,1
3	Уменьшение ошибки оценки дисконтированной доли за счёт учёта тренда инфляции Reducing the proportion of the discounted valuation error due account of the inflation trend	7,3
4	Уменьшение ошибки оценки дисконтированной доли за счёт оптимизации экономической информации Reducing the proportion of the discounted estimation error due to the optimization of the economic information	2,1
5	Уменьшение ошибки оценки дисконтированной доли за счёт сбора дополнительной экономической информации Reducing the error estimation discounted the proportion of at the expense of collecting additional economic information	0,7

Заключение

Таким образом, нейросетевые модели земельного рынка, характеризующиеся возможностью нелинейного описания сложных динамических систем, адекватны современной рыночной ситуации. Однако вследствие неполноты и значительной кластеризации эмпирических данных для земельного рынка и рынка строительной продукции примеры успешного применения нейросетевых технологий к таким системам весьма немногочисленны. Практическая реализация нейросетевых моделей становится возможной только после

формирования репрезентативного набора обучающих и тестовых примеров. Эта операция может быть выполнена альтернативными методами: либо а основе растяжения области сгущения примеров, либо за счёт сбора дополнительной информации в области их малой плотности. Однако второй метод весьма затратен и должен быть использован только после исчерпания возможностей первого. Экономические результаты уменьшения ошибки оценки земельной составляющей при использовании разработанных в этой работе алгоритмов свидетельствуют об их высокой эффективности по соотношению затраты/отдача.

ЛИТЕРАТУРА

- 1 Грешилов А.А. Математические методы принятия решений, М.: МГТУ им. Н.Э. Баумана, 2006, 583 с.
- 2 Stanley J. Introduction to Neural Networks. Sierra Madre, CA // California Scientific Software. 2011. 311 p.
- 3 Hinton. G.E., Sejnowski. T.J. Learning and Relearning in Boltzmann Machines // P-rullel Distributed Proceedings. 1996. V. 11. P. 282–317.
- 4 Цыпкин А.И. Нейронные сети: История развития теории. М.: ИПРЖР, 2011, 517 с.
- 5 Бир С. Мозг фирмы. М.: Радио и связь, 1993. 524 с.
- 6 Eldon Y. L., San-Hsing, Ming-Hsiung, Chia-Yi Applications Artificial neural networks and their business applications // Information & Management. Institute of Information Management, National Chung Cheng University of Taiwan, R.O.C. 2014, P. 303–313.
- 7 Фролов Ю.В., Пастухов Е.С. Мониторинг изменений в банковском сообществе России с применением самоорганизующихся карт Кохонена // Банковские технологии. 2012. № 11.
- 8 Недосекин А.О., Заболоцкий С.Н. Подход к учету долговых обязательств в программах фондового менеджмента // Аудит и финансовый анализ. 2001. № 1. 31 с.
- 9 Карпович М.А. Нейросетевые методы оценки затрат на научно исследовательские и проектно-изыскательские работы // Вестник Воронежского государственного университета инженерных технологий». 2014. № 1. С. 235–240.
- 10 Галушкин А.И. Применения нейрокомпьютеров в финансовой деятельности. Новосибирск: Наука, 2002. 215 с.
- 11 Barron A.R. Predicted Squarer Errors: A Criterion for Automatic Model Selection. New York: Marcel Dekker, 2006. 311 p.

REFERENCES

- 1 Greshilov A.A. Matematicheskie metody prinyatiia reshenii [Mathematical methods of decision-making] Moscow, MGТУ im. N.E. Bauman, 2006. 583 p. (in Russian)
- 2 Stanley J. Introduction to Neural Networks. Sierra Madre, CA. California Scientific Software, 2011, 311 p.
- 3 Hinton G.E., Sejnowski. T.J. Learning and Relearning in Boltzmann Machines». P-rullel Distributed Proceedings, 1996, vol. 11, pp. 282–317.
- 4 Tsypkin A.I. Neironnye seti: Istoriia razvitiia teori [Neural Networks: Theory of History] Moscow, IPRZhR, 2011, 517 p. (in Russian)
- 5 Bir S. Mozg firmy [Firms brain] Moscow, Radio i sviaz', 1993. 524 p. (in Russian)
- 6 Eldon Y. L., San-Hsing, Ming-Hsiung, Chia-Yi Applications Artificial neural networks and their business applications. Information & Management. Institute of Information Management, National Chung Cheng University of Taiwan, R.O.C., 2014, pp. 303–313.
- 7 Frolov Yu.V., Pastukhov E.S. Monitoring changes in the Russian banking community with the use of self-organizing maps of Kohonen. *Bankovskie tekhnologii* [Banking technologies] 2012, no. 11. (in Russian)
- 8 Nedosekin A.O., Zabolotskiy S.N. The approach to recording debt stock management programs. *Audit i finansovyi analiz* [Audit and Financial Analysis] 2001, no. 1, 31 p. (in Russian)
- 9 Karpovich M.A. Neural network methods of estimating costs for scientific research, design and development works *Vestnik Voronezhskogo gosudarstvennogo universiteta inzhenernykh tekhnologiy* [Proceedings of VSUET] 2014, no.1, pp. 235–240. (In Russian)
- 10 Galushkin A. I. Primenenie neirokomputerov v finansovoi deiatelnosti [Applications of neuro-computers in financial activities] Novosibirsk, Nauka, 2002, 215p. (in Russian)
- 11 Barron A.R. Predicted Squarer Errors: A Criterion for Automatic Model Selection. New York: Marcel Dekker, 2006. 311 p.

СВЕДЕНИЯ ОБ АВТОРАХ

Мирон А. Карпович доцент, кафедра экономики и основ предпринимательства, Воронеж. гос. архитектурно-строительный ун-т, ул. 20-летия Октября, 64, г. Воронеж, 394006, Россия, karpovich.am@gmail.com

Екатерина Г. Фролова аспирант, кафедра экономики и основ предпринимательства, Воронеж. гос. архитектурно-строительный ун-т, ул. 20-летия Октября, 64, г. Воронеж, 394006, Россия, eg.frolova@gkp.vrn

КРИТЕРИЙ АВТОРСТВА

Екатерина Г. Фролова предложила методику проведения эксперимента и организовала производственные испытания

Мирон А. Карпович написал рукопись, корректировал её до подачи в редакцию и несёт ответственность за плагиат

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ПОСТУПИЛА 21.03.2016

ПРИНЯТА В ПЕЧАТЬ 18.04.2016

INFORMATION ABOUT AUTHORS

Myron A. Karpovich Associate Professor, Department of Economics and the basics of entrepreneurship, Voronezh state Architecture and Construction Univ Str. October 20 anniversary, 64, Voronezh, 394006, Russia, karpovich.am@gmail.com

Ekaterina G. Frolova a graduate student, economics and the basics of entrepreneurship department, Voronezh. state. Architecture and Construction University, 20 Otyabya str, 64, Voronezh, 394006, Russia, eg.frolova@gkp.vrn.ru

CONTRIBUTION

Ekaterina G. Frolova proposed a methodology of experimental swipe-organized and production tests

Myron A. Karpovich wrote the manuscript, correct-ment it before filing in the office and is liable for plagiarism

CONFLICT OF INTEREST

The authors declare no conflict of interest.

RECEIVED 3.21.2016

ACCEPTED 4.18.2016