

УДК 69:001.89

Доцент М.А. Карпович

(Воронежский государственный архитектурно-строительный университет)
кафедра экономики и основ предпринимательства. тел. (473) 271-54-00
E-mail: karpovich.am@gmail.com

Associate Professor M.A. Karpovich

(Voronezh state university of architecture and civil engineering). Department of economy and the basics of entrepreneurship. phone (473) 271-54-00
E-mail: karpovich.am@gmail.com

Нейросетевые методы оценки затрат на научно-исследовательские и проектно-изыскательские работы

Neural network methods for estimating the costs of research and design and survey works

Реферат. В статье рассмотрены нейросетевые методы оценки затрат на научно-исследовательские и проектно-изыскательские работы при строительстве автомобильных дорог, позволяющие на основе данных по запроектированным ранее объектам произвести ранжирование ценообразующих факторов по степени их влияния на цену научно-исследовательских и проектно-изыскательских работ, выполняемых для проектов государственно-частного партнерства (ГЧП). Преимущества нейросетевых методов определяется следующими обстоятельствами: нейросетевые модели автоматически учитывают взаимное влияние ценообразующих факторов; нейросетевые методы полностью свободны от субъективных факторов. Оптимизация нейросети позволила ранжировать ценообразующие параметры по степени их влияния на цену научно-исследовательских и проектно-изыскательских работы в условиях ГЧП. При этом 4 параметра: «Статус заказчика», «Вид работ», «Вид конкурса» и «Категория дороги» в совокупности более чем на 87 % определяют удельную цену проекта. Конкретные расчеты свидетельствуют о том, что нейросети позволяют весьма точно (с относительной ошибкой меньше 0.2 %) описывать большую часть объектов, и лишь небольшую часть – менее 5 % подмножеств, со значительной ошибкой - от 9 % до 17 %.

Summary. The article describes the neural network methods for estimating the costs of research and design and survey work in the construction of roads, allowing on the basis of data previously projected objects produce price-setting ranking factors according to their impact on the cost of research and development works undertaken for public-private partnership (PPP). Advantages of neural network methods is determined by the following circumstances : neural network models automatically take into account the mutual influence of the pricing factors ; Neural methods are completely free of subjective factors. Optimization of neural network allowed rank price-setting parameters according to their impact on the cost of research , design and survey work under PPP . This causes the 4 "customer status", "Type of work", "Kind of competition" and "Road category" in the aggregate more than 87 % determined by the unit price of the project. Specific calculations show that the neural network allow very accurate (with a relative error less than 0.2%) to describe most of the objects , and only a small fraction - less than 5 % of subsets , with significant error - from 9 % to 17 %.

Ключевые слова: нейросетевые методы, государственно-частное партнерство, научно-исследовательские работы, проектно-изыскательские работы.

Keywords: neural network methods, public-private partnerships, research and development, design and survey works.

Существующие методы определения стоимости проектных работ на строительство (реконструкцию) автомобильных дорог и сооружений на них [1] предполагают весьма трудоемкую процедуру расчета, требующую сбора значительного объема информации. Расчеты основаны на большом числе коэффициентов, описывающих экономическое состояние района проектирования и технико-экономические характеристики проекта и нуждающихся в постоянном мониторинге и изменении. Вследствие этого традиционные способы расчета предельных цен

на научно-исследовательские и проектно-изыскательские работы в условиях ГЧП свидетельствует об их ограниченной применимости в условиях динамичного рынка, что связано со следующими обстоятельствами:

- 1) указанные методы не учитывают синергетическое или антагонистическое влияние ценообразующих факторов;
- 2) эти методы неизбежно имеют значительную субъективную составляющую (например, при определении значений повышающих или понижающих коэффициентов);

3) указанные методы требуют для своего применения значительных, часто практически недоступных, объемов информации;

4) традиционные методы плохо отражают реалии быстро развивающихся рынков, особенно в случае, когда накопление количественных изменений к качественным переменам;

5) эти методы требуют значительных затрат сил и времени и высокой квалификации на всех этапах их применения.

Свободный от этих недостатков метод методов расчета начальной (максимальной) цены контракта на проектирование элементов транспортной инфраструктуры на принципах ГЧП может быть основан на нейросетевых технологиях. Преимущества нейросетевых методов определяется следующими обстоятельствами:

1) нейросетевые модели автоматически учитывают взаимное влияние ценообразующих факторов;

2) нейросетевые методы полностью свободны от субъективных факторов: даже если на стадии определения архитектуры нейронной сети в ее структуру будут включены параметры, неадекватные поставленной задаче, генетический алгоритм автоматически отфильтрует их;

3) информация, необходимая для тренировки сети, является легко доступной;

4) адаптация нейронной сети к быстро меняющейся ситуации происходит без изменения ее конфигурации лишь путем тренировки имеющейся сети на новых практических примерах. При этом сети Кохонена позволяют в автоматическом режиме выявлять качественно новые явления;

5) квалифицированные специалисты необходимы лишь на стадии определения первоначальной архитектуры сети и изменения ее конфигурации при возникновении качественных изменений рыночной ситуации. На остальных этапах эксплуатация нейросети осуществляется техническими работниками.

Эти свойства нейросетевых технологий позволяют использовать их в условиях реального быстро меняющегося рынка и взрыв интереса к нейронным сетям. Они находят успешное применение в самых различных областях – бизнесе, медицине, технике, геологии, физике, экономике и др. В частности, нейросетевые методы позволяют успешно решать задачи разработки алгоритмов нахождения аналитического описания закономерностей функционирования экономических объектов (предприятие, отрасль, регион) [2].

Эти алгоритмы применяются также и к прогнозированию показателей объектов. Нейронные сети нелинейные по своей природе и представляют собой исключительно мощный

метод моделирования, позволяющий воспроизводить чрезвычайно сложные зависимости. Наиболее эффективно нейронная сеть используется тогда, когда неизвестен точный вид связей между входами и выходами, - если бы он был известен, то связь можно было бы моделировать непосредственно.

Структура нейросетевой модели позволяет разделить этап конфигурирования сети и этап ее эксплуатации. При этом первый этап требует значительных и разнообразных знаний, как в предметной области, так и математических методов. В противоположность этапу конфигурирования, на этапе эксплуатации уровень знаний, необходимый для успешного применения нейронных сетей, гораздо скромнее, чем, например, при использовании традиционных методов статистики. Эта особенность нейронных сетей позволяет осуществить их практическое использование.

Для решения задачи нейросетевыми методами необходимо собрать обучающий набор данных, представляющих собой ряд наблюдений, для которых указаны значения входных и выходных переменных. На первом этапе нужно решить, какие переменные использовать, сколько и каких наблюдений собрать. Ответ на этот вопрос основывается на знании предметной области. На первом этапе используются все переменные, которые могут влиять на результат, а на последующих этапах необходимо сократить это множество. При этом для эффективной работы нейронной сети важно не только количество примеров, но сбалансированность их набора. Множество примеров должно равномерно покрывать пространство событий. В противном случае правильные выводы сеть будет делать только в окрестностях точек сгущения набора примеров.

Нейронная сеть принимает на входе числовые значения и выдает на выходе также числовые значения. При этом алгоритмы ее работы построены так, что входные и выходные значения всегда принадлежат некоторой ограниченной области. Поэтому при решении реальных задач методами нейронных сетей вся информация должна быть представлена в числовом виде и необходимы этапы предварительной обработки - преобработки - и заключительной обработки пост-обработки данных. Числовые значения должны быть приведены в масштаб, подходящий для сети. Номинальные переменные, принимающие значения из фиксированного списка, преобразуются в числовую форму. Тип выходной переменной определяется задачей, решаемой нейронной сетью. В задачах классификации выходная переменная

является номинальной, а в задачах регрессии – числовой (например, действительное значение цены). Соответствующими способами и решаются задачи пост-процессирования данных.

Специфичность экономических задач проявляется не только в способах их синтеза, но и характере делаемых приближений (и связанных с ними ошибок). Отличия в поведении системы и ее информационной модели возникают вследствие свойств эмпирических данных. Такие особенности в данных и в постановке задач требуют особого отношения к ошибкам экономических моделей, проявляющиеся, в частности, в разделении ошибка обучения и ошибка обобщения.

Прямое измерение указанной ошибки модели на практике не достижимо, поскольку системная функция при произвольных значениях аргумента не известна. Однако возможно получение ее оценки. При использовании базы данных наблюдений за системой, для обучения может отводиться некоторая ее часть, называемая в этом случае обучающей выборкой. Для обучающих примеров отклики системы известны. Норма невязки модельной функции и системной функции на множестве играет важную роль в информационном моделировании и называется ошибкой обучения модели.

Оценка ошибки обобщения является принципиальным моментом при построении информационной модели. На первый взгляд может показаться, что сознательное не использование части примеров при обучении может только ухудшить итоговую модель. Однако без этапа тестирования единственной оценкой качества модели будет лишь ошибка обучения, которая, как уже отмечалось, мало связана с предсказательными способностями модели. В профессиональных исследованиях могут использоваться несколько независимых тестовых выборок, этапы обучения и тестирования повторяются многократно с вариацией начального распределения весов нейросети, ее топологии и параметров обучения. Окончательный выбор "наилучшей" нейросети выполняется с учетом имеющегося объема и качества данных, специфики задачи, с целью минимизации риска большой ошибки обобщения при эксплуатации модели.

Эта цель достигается на следующем этапе конструирования сети. На этом этапе необходимо определить число нейронных слоев и число элементов в каждом из них. Для сети Кохонена ответ на эти вопросы определяется ее строением: она всегда содержит два слоя – входной и выходной, называемый также слоем топологической карты. Для многослойного персептрона

формально ответ на первый вопрос дается теоремой Колмогорова [3], согласно которой сколь угодно сложная зависимость моделируется многослойным персептроном с двумя промежуточными слоями. Однако практически оптимальная сложность сети определяется сложностью связей, которые она отображает.

Анализ особенностей функционирования и требований, предъявляемых к нейросетевой модели (НСМ), описывающей сложные социально-экономические системы, позволил определить следующий алгоритм ее формирования, тестирования и эксплуатации:

1. Первым этапом построения НСМ является формулирование содержательной модели экономической системы на базе анализа ее реального функционирования, что позволяет определить количественные и качественные характеристики входного и выходного слоев нейронной сети.

2. Далее определяется объем, энтропия и степень достоверности реально достижимой рыночной информации. Анализируется и прогнозируется вклад различных областей конфигурационного пространства событий в реальные экономические процессы.

3. На основе этого анализа из полного массива статистической информации отбирается набор примеров, наиболее полно и всесторонне отражающий экономические реалии.

4. Анализируется возможность обучения нейронной сети с определенными на первом этапе характеристиками входного и выходного слоев на доступном наборе примеров. Анализ должен, в первую очередь, основываться на необходимости учета проклятия размерности и переобучения сети.

5. Параметры задачи анализируются сетями Кохонена. В результате такого анализа строится топологическая карта задачи, данные кластеризуются, полученные кластеры ранжируются по степени влияния на окончательное решение. Кроме того, возможно выявление социально-экономической сути кластеров.

6. В случае отсутствия достаточной информации или при наличии ее избытка необходимо возвратиться к п. 1 и переконфигурировать входной и выходной слои нейронной сети для приведения ее сложности в соответствие с содержательной моделью экономической системы и качеством и количеством доступной информации.

7. Пп. 1-5 итерационно повторяются до достижения самосогласованности модели, после чего возможен переход к последующим шагам алгоритма.

8. Из набора входных переменных выделяются номинальные – то есть принимающие конечное число значения из фиксированного набора (например – статус заказчика может принимать значения из набора ({«федеральный», «областной», «муниципальный»}) и числовые.

9. Номинальные переменные разделяются на ранжируемые (кодируются во входном наборе монотонной последовательностью натуральных чисел) и неранжируемые (кодируются набором векторов).

10. Выбирается метод пре-процессирования, позволяющий отобразить числовые переменные в фиксированную ограниченную область изменения и выполняется это отображение.

11. Выбирается метод кодировки выходного слоя для задач классификации в виде номинальных переменных, а для задач оценки – в виде числовых переменных. При этом в последнем случае выполняется пост-процессирование, отображающее вектор выходных свойств из фиксированного диапазона в реально наблюдаемый, соответствующий содержательной концепции экономического и производственного процесса.

12. Определяется связность и архитектура сети.

13. Определяется количество промежуточных нейронных слоев и их параметры. При этом, хотя в соответствии с теоремой Колмогорова сколь угодно сложная зависимость может быть отображена двухслойной сетью, в некоторых практических случаях рациональными оказываются и более сложные сети с соответствующим уменьшением числа нейронов в каждом слое.

14. В соответствии с содержательной экономической концепцией определяются методы контроля и диагностики сети.

15. Определяются методы тренировки сети и выполняется реализация выбранного алгоритма.

16. На сформулированном в пп. 2-1 эмпирическом массиве осуществляется обучение сети и в соответствии с п. 13 осуществляется контроль качества ее работы.

17. С помощью генетического алгоритма оптимизируется входной слой сети.

18. Модифицируются число и состав промежуточных слоев и пп. 16-18 итерационно повторяются до достижения минимальной сложности нейросети, удовлетворяющей требованиям практики.

19. При достижении достаточной для практических применений точности описания экономической системы возможен переход к этапу эксплуатации сети, сводящемуся к

пополнению набора примеров, анализу полученного эмпирического массива в соответствии с требованиями п. 3 и обучению сети на расширенном массиве.

20. Расширенный массив в соответствии с п. 5 анализируется сетью Кохонена и при отсутствии новых классов данных результаты нейросетевой оценки используется в практике. При возникновении новых классов эмпирических данных необходимо вернуться к п. 4.

В качестве входных переменных используется следующий набор ценообразующих параметров:

1. Статус заказчика {федеральный, областной, муниципальный}.
2. Вид работ {строительство, реконструкция, капитальный ремонт, ремонт}.
3. Вид конкурса {открытый, субподряд}.
4. Уровень экспертизы проекта {государственная, областная, нет}.
5. Категория дороги.
6. Число полос движения.
7. Дата начала проекта.
8. Длительность выполнения проекта.
9. Стоимость проекта в текущих ценах.
10. Протяженность моста.
11. Площадь моста.
12. Авторский надзор {да, нет}.
13. Удельная цена на единицу площади.
14. Погонная цена на единицу длины.
15. Наименование проекта.

Здесь курсивом отмечены номинальные ранжируемые переменные; полужирным шрифтом – переменные выходного слоя; полужирным курсивом – неучитываемая информационная переменная.

Последняя неучитываемая переменная при тренировке может служить только для взаимодействия модели с ее пользователями и интерпретации результатов работы нейросети. Переменные 7-12 не являются независимыми. Так, в частности, удельная и погонная цены должны быть приведены к унифицированному временному промежутку (в работе - на 01.01.2013), для чего необходимо учесть дисконтирование, являющееся универсальной методикой приведения денежных потоков к фиксированному моменту времени, основанной на понятиях сложных процентов. Формула дисконтированной стоимости денежных потоков в моменты времени, отстоящие на N промежутков, выглядит следующим образом:

$$P_N = P_0(1 + r)^N, \quad (1)$$

где r - ставка дисконтирования за промежуток дискретизации, в данной работе принятый равным 1 месяцу. Поскольку в экономике ча-

ще задаются годовые ставки, необходимо определить связь между ставками дисконтирования за кратные периоды (r_Y и r_M - годовая и месячная соответственно), задаваемую соотношением:

$$1 + r_Y = (1 + r_M)^{12}, \quad (2)$$

решение которого имеет вид:

$$r_M = \exp\left[\frac{1}{12} \ln(1 + r_Y)\right] - 1. \quad (3)$$

Так, например, при годовой инфляции в 15 %, месячная ставка дисконтирования принимает значение

$$r_M = \exp\left(\frac{\ln 1.15}{12}\right) - 1 = 0.0117. \text{ Для опреде-}$$

ленности при дисконтировании приняты следующие схемы расчетов:

1. Для проектов, длительность выполнения которых не превышает двух кварталов, считается, что 25 % платежей приходится на начало срока проектирования, а 75 % - на его конец. Применение этой схемы приводит к следующему соотношению:

$$P_N = P_0 \left[\frac{(1+r)^{N+\Delta N}}{4} + \frac{3(1+r)^N}{4} \right] = \frac{P_0(1+r)^N}{4} [(1+r)^{\Delta N} + 3] \quad (4)$$

2. Для проектов, длительность выполнения которых превышает два квартала, оплата выполненных работ выполняется равномерно поквартально. В этой схеме

$$P_N = \frac{P_0(1+r)^N}{3N} [1 + 1 + r + (1+r)^2] = \frac{P_0(3 + 3r + r^2)(1+r)^N}{3N} = \frac{P_0(1+r)^N}{N} \left(1 + r + \frac{r^2}{3}\right) \quad (5)$$

Здесь N – промежуток (в месяцах) от окончания проектирования до фиксированной даты, ΔN – длительность проектирования в месяцах. Поскольку длительность периода дискретизации времени зафиксирована, индекс «М» у ставки дисконтирования здесь и далее опущен.

Реализация нейросети в данной работе выполнена в лицензионном пакете STATISTICA Neural Networks (версия реализации 2.3). Результат анализа топологической карты сети Кохонена свидетельствует о наличии четко выделенных кластеров, определяемых переменными «2. Вид работ» и «3. Вид конкурса». В эти кластеры входит более 81 % исходных примеров, что свидетельствует о детерминированности процессов формирования статистического массива. Таким образом, анализ имеющейся информации, степени ее

кластеризации, полноты и связанности позволяет перейти к следующему этапу построения нейросети – формированию входного слоя.

В имеющемся наборе ценообразующих параметров выделяются номинальные переменные - «Статус заказчика», «Вид работ», «Вид конкурса», «Авторский надзор». Все они являются ранжируемыми и, следовательно должны описываться в скалярной форме.

Переменные «5. Категория дороги», «6. Число полос движения», «7. Дата начала проекта», «8. Длительность проекта», «9. Стоимость проекта в текущих ценах», «10. Протяженность моста», «11. Площадь моста», «13. Удельная цена на единицу площади», «14. Погонная цена на единицу длины» принимают действительные значения. При этом переменные №№ 5 и 6 принимают значения из фиксированного набора и формально могут быть описаны так же и как номинальные. Переменные №№ 13 и 14, в зависимости от экономического содержания задачи, могут быть отнесены как к входному, так и к выходному слою нейросети.

Анализ дисперсии тренировочного множества свидетельствует об отсутствии резкой неоднородности распределения примеров в пространстве событий по всем входным параметрам, кроме номинальных переменных «Статус заказчика» и «Вид работ», что позволяет выполнить линейное преобразование большинства переменных. Для указанных входных переменных избежать насыщения сигмоидной функции можно, используя логарифмическое преобразование вида

$$\ln X = \ln X_{\min} + c \cdot (\ln X_{\max} - \ln X_{\min}), \quad (6)$$

позволяющее растянуть область малых значений номинального ценообразующего параметра.

Свойства функции активации нейрона и экономического анализа предопределил использованный в данной работе метод преобразования исходных данных:

1. Во-первых, для номинальных переменных интервал для любого множества одинаков, следствие чего $[x_{\min}, x_{\max}] = [-1, 1]$.

2. Во-вторых, для двух числовых переменных («Категория дороги» и «Число полос движения») существуют принципиальные как нижняя, так и верхняя границы, вследствие чего для них интервал преобразования совпадает с интервалом номинальных переменных.

3. И, в-третьих, в условиях недостаточной информации и нестационарного рынка далекая экстраполяция является необоснованной, вследствие чего для остальных числовых переменных приняты осторожные значения

границ $[x_{\min}, x_{\max}] = [-0.9, 0.9]$ близкие к границам интервала $[-1, 1]$.

4. Исключение сделано для параметра «Длительность выполнения проекта», для которого существует принципиальная нижняя граница, и вклад которого в регрессионные зависимости незначителен. В этом случае принят интервал $[x_{\min}, x_{\max}] = [-1, 0.5]$, обеспечивающий возможность экстраполяции результатов на область значительного увеличения длительности проектирования для сложных и уникальных объектов.

В результате обучения нейронная сеть формулирует правила определения значения цены по значениям ценообразующих параметров. Ограниченный объем доступной входной информации делает недостаточными стандартные методы диагностики обучаемости и степени обобщения нейросети. Поэтому в данной работе сформулированы дополнительные критерии качества принятых системой решений. Выбор оптимальной архитектуры сети проводился по следующим четырем критериям:

- 1) минимальное значение δ ;
- 2) минимальное количество наблюдений, при описании которых относительная ошибка регрессии превышает фиксированное значение;
- 3) максимальная ошибка отдельного наблюдения обучающего множества;
- 4) максимальное число наблюдений, описываемых с относительной ошибкой, меньшей фиксированного значения.

Первоначальная оптимизация структуры сети состоит в использовании сетей Кохонена, которая была выполнена на первоначальном этапе ее конфигурирования. Известны

альтернативные сетям Кохонена методы понижения для размерности решаемой задачи.

Следующий шаг алгоритма состоит в оптимизации структуры сети, то есть выборе числа слоёв, числа нейронов и числа связей для каждого нейрона. Эта часть процедуры реализации нейросетевой технологии является, пожалуй, наиболее сложной проблемой. Общий замкнутый алгоритм этой процедуры в настоящее время не построен. Поэтому на практике применяются различные эмпирические стратегии поиска оптимальной структуры сети:

- 1) постепенное наращивание;
- 2) построение заведомо слишком сложной сети с последующим упрощением;
- 3) поочерёдное наращивание и упрощение.

После выполнения всех шагов алгоритма, соответствующих структурированию нейросети, был выполнен переход к этапу эксплуатации, тестирования и модернизации.

Оптимизация входного слоя нейросети позволила ранжировать ценообразующие параметры по степени их влияния на цену научно-исследовательских и проектно-изыскательских работ в условиях ГЧП. При этом 4 параметра: «Статус заказчика», «Вид работ», «Вид конкурса», «Категория дороги» в совокупности более чем на 87 % определяют удельную цену. Ввиду синергетического вклада этих параметров, доля каждого из них не может быть выделена. Конкретные расчеты свидетельствуют о том, что нейросети позволяют весьма точно (с относительной ошибкой меньше 0.2 %) описывать большую часть объектов, и лишь небольшую часть – менее 5 % подмножеств, со значительной ошибкой - от 9 % до 17 %.

ЛИТЕРАТУРА

1 Гасилов В.В., Карпович М.А., Замчалова С.С. и др. Методические рекомендации по определению стоимости проектных работ на строительство (реконструкцию) автомобильных дорог и сооружений на них. М.: Росавтотдор, 2003.

2 Матвеев М.Г. Модели и методы искусственного интеллекта. Применение в экономике. М.: Финансы и статистика; ИНФРА – М, 2008. 448 с.

3 Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Наука, 2009. 487 с.

1 Gasilov V.V., Karpovich M.A., Zamchalova S.S. et al. Metodicheskie rekomendatsii po opredeleniiu stoimosti proektnykh rabot na stroital'stvo (rekonstruksiiu) avtomobil'nykh dorog i sooruzhenii na nikh [Methodical recommendations on how to determine the cost of design work for construction (reconstruction) of roads and structures on them]. Moscow, Rosavtodor, 2003.

2. Matveev M.G. Model ii metody iskusstvennogo intellekta. Primenenie v ekonomike [Models and methods of artificial intelligence. Applications in economics]. Moscow, Finansy i statistika, INFRA - M, 2008. 448 p.

3 Kolmogorov A.N., Fomin S.V. Elements of theory of functions and functional analysis. Moscow, Nauka, 2009. 487 p.

REFERENCES