

УДК 004.891.2

ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОЦЕНКИ СТОИМОСТИ ПОДЕРЖАННЫХ АВТОМОБИЛЕЙ

Гамагина А.А., Гуда А.Н., Цуриков А.Н.

ФГБОУ ВО «Ростовский государственный университет путей сообщения», Ростов-на-Дону,
e-mail: tsurik7@yandex.ru

Статья посвящена разработке экспертной системы оценки стоимости подержанных автомобилей. В качестве целевой марки и модели автомобиля выбран «Rapid» известного чешского производителя «Skoda». В работе осуществлена математическая постановка задачи определения стоимости автомобиля в виде задачи классификации. Для решения задачи предложено использовать технологию искусственных нейронных сетей. В качестве искусственной нейронной сети, подлежащей обучению, использован универсальный классификатор типа «многослойный перцептрон» с двумя скрытыми слоями. В качестве средства моделирования нейронной сети применяется программа-нейропакет «NeuroSolutions». Определены входные и выходные параметры обучаемой искусственной нейронной сети. Сбор статистики для обучения искусственной нейронной сети осуществлен на российских интернет-сайтах, специализирующихся на продаже и покупке подержанных автомобилей, в частности на auto.ru. В итоге созданная система способна решить представленную задачу с достаточной степенью точности. Создан файл динамически подключаемой библиотеки DLL, содержащий искусственную нейронную сеть, прошедшую процесс обучения. Для взаимодействия с DLL-модулем разработано приложение на языке C++ (в Visual Studio) с интерфейсом пользователя. В статье рассмотрены перспективы внедрения результатов исследования. Проанализированы пути повышения качества работы созданной системы. Система может применяться для сервис-ориентированной оценки автомобилей.

Ключевые слова: задача классификации, искусственная нейронная сеть, многослойный перцептрон, обучение нейронной сети, оценка стоимости, подержанный автомобиль, «Шкода Рапид», экспертная система

THE APPLICATION OF ARTIFICIAL NEURAL NETWORKS FOR VALUATION OF USED CARS

Gamagina A.A., Guda A.N., Tsurikov A.N.

Rostov State Transport University, Rostov-on-Don, e-mail: tsurik7@yandex.ru

The article is devoted to the development of an expert system of valuation of used cars. As the target brand and model of the car chosen «Rapid» known Czech manufacturer «Skoda». The paper presents a mathematical formulation of the problem of determining the value of the car in the form of the classification problem. The technology of artificial neural networks is proposed to use for solve the problem. As an artificial neural network to be trained, a universal classifier of the «multilayer perceptron» type with two hidden layers was used. The «NeuroSolutions» neuropackage program is used as the neural network modeling tool. The input and output parameters of the artificial neural network are determined. Collection of statistics for the training of artificial neural network carried out on the Russian Internet sites specializing in the sale and purchase of used cars, in particular auto.ru. As a result, the created system is able to solve the presented problem with a sufficient degree of accuracy. A dynamic-link library (DLL) module is generated that contains a trained neural network. To work with the DLL-module was written software application in C++ (Visual Studio) with a user interface. The article discusses the prospects of implementing the results of the study. The ways of improving the quality of the created system are analyzed. The system can be used for service-oriented evaluation of cars.

Keywords: classification problem, artificial neural network, multilayer perceptron, training of neural network, valuation, used car, «Skoda Rapid», expert system

Технология искусственных нейронных сетей (ИНС) в наши дни получила широкое распространение [1, 2]. Практически ежедневно поступают сообщения о новых областях успешного применения ИНС, создания программных продуктов и устройств на их основе.

В связи с этим большой интерес представляет изучение возможностей применения ИНС и машинного обучения при создании современных сервис-ориентированных программных продуктов, особенно если решаемые ими задачи можно свести к решению задач классификации, с которыми, как известно, ИНС умеют успешно справляться [3, 4].

Одной из них является задача оценки подержанных автомобилей, а именно – определение цены конкретного экземпляра

транспортного средства. Цена – количество денег, в обмен на которые продавец согласен продать товар [5].

Об использовании ИНС в области определения цен известно, но аналогичные сервисы (с использованием элементов машинного обучения) пока немногочисленны и находятся в стадии разработки или тестового запуска [6]. Тема имеет важность в связи с расширением количества интернет-площадок, увеличением доли сделок с бывшими в употреблении авто, снижением спроса на новые машины из-за сложной ситуации в экономике.

Цель исследования: изучение возможности применения искусственной нейросети в программных сервисах, предназначенных для приблизительной оценки цены поде-

ржанных автомобилей отдельно выбранной марки и модели по ряду характеристик.

Материалы и методы исследования

Для проведения исследования необходимо: определить марку и модель автомобиля, подобрать источник данных для обучения ИНС, выбрать средства моделирования и обучения ИНС, осуществить математическую постановку задачи, определить архитектуру ИНС, количество слоев и нейронов. Эти и другие пункты будут раскрыты ниже подробнее.

Выбор модели и марки автомобиля

Решено выбрать в качестве целевой марки и модели – «Рапид» (*Rapid*) чешского производителя «Шкода» (*Skoda*), входящего в концерн «Фольксваген» (*Volkswagen*).

Выбор обусловлен рядом причин. Модель довольно популярна на рынке, появилась относительно недавно (2012–2014 гг.), но при этом выпускается достаточное время, чтобы на площадках по продаже авто могло появиться достаточное количество объявлений.

Ранее автомобили с таким названием не производились (модели 1935 и 1984 г. можно не учитывать, на рынке их практически нет). За время производства (6 лет) *Rapid* не происходило смены поколений, был проведен лишь ряд рестайлингов, существенно не повлиявших на внешний вид и техническое оснащение автомобиля [7].

Автомобили, собираемые в Чехии, России, Украине и Казахстане, почти не имеют отличий. Кузов автомобиля представлен практически единственным вариантом – лифтбек.

Технически все автомобили имеют три варианта коробок передач (механическая, автоматическая, роботизированная *DSG*) и двигателей (атмосферные – объем 1,6 л (90 л.с. или 110 л.с.), а также турбированный – объем 1,4 л, 125 л.с.). Ранее выпускались автомобили с двигателями 105 и 122 л.с., которые присутствуют на вторичном рынке [8].

Таким образом, *Rapid* – идеальный вариант для классификации на начальном этапе применения ИНС, поэтому данный автомобиль и будет использоваться в нашей работе.

Источник данных для обучения ИНС

В настоящее время в интернете можно обнаружить большое количество сайтов, специализирующихся на продаже и покупке подержанных автомобилей. Нами

были изучены следующие ресурсы: *auto.ru*, *drom.ru*, *car.ru*, *genser.ru* и автомобильный раздел сайта *avito.ru*.

В итоге остановим свое внимание на старейшем и крупнейшем в России сайте *auto.ru* и на портале *drom.ru*. Статистика по автомобилям исследуемой модели, характеристикам и связанной с ними цене собрана с этих сайтов.

На *auto.ru* размещено свыше 500000 объявлений о продаже подержанных автомобилей. При этом объявлений о продаже *Skoda Rapid* – около 3 тысяч по всем регионам.

Сбор статистики осуществлялся с сайтов в ручном режиме в течение одной недели в ноябре 2017 г. среди объявлений, опубликованных в ряде географически близких регионов на юге России (Ростов-на-Дону + 500 км). Результаты обрабатывались с использованием табличного процессора *Microsoft Excel*.

Средства моделирования и обучения ИНС

В качестве средства моделирования ИНС выбрано использование программы-нейропакета – программного продукта, позволяющего создавать и обучать ИНС, эмулируя их поведение на обычном персональном компьютере.

В качестве альтернатив были рассмотрены инструменты: *NeuroSolutions*, *Process Advisor*, *NeuralWorks Professional II*, *BrainMaker Pro*, *NeuroShell*, *MatLab + Neural Network Toolbox*.

Каждый из них обладает своими достоинствами и недостатками, которые подробно описаны в статьях [9, 10]. Для работы был выбран *NeuroSolutions* компании *NeuroDimension*.

Математическая постановка задачи

Математическая постановка задачи определения стоимости практически аналогична постановке задачи классификации [3]. Задача формулируется как задача классификации, где $G = [\text{пробег, возраст, мощность двигателя, тип коробки передач}]$ – множество параметров отдельно взятого автомобиля, $Q = \{A, B, C, D, E, F\}$ – диапазон изменения цен. Есть некое отображение $g^*: G \rightarrow Q$, его значения известны на примерах выборки $G_j = \{(g_1, q_1), \dots, (g_p, q_p)\}$. Надо создать систему $\beta: G \rightarrow Q$, которая сможет верно классифицировать любой экземпляр автомобиля $g \in G$.

Определение входных и выходных параметров

Вспользуемся сайтом по продаже подержанных автомобилей *auto.ru* (рис. 1).



Рис. 1. Объявление о продаже *Skoda Rapid* на сайте *auto.ru*

Можно заметить, что цена главным образом определяется по четырем основным параметрам: пробег, мощность двигателя, тип коробки передач, год выпуска (возраст автомобиля). Их мы и выберем в качестве входов ИНС.

Определим следующие диапазоны изменения входных переменных. Пробег – от 100 до 750000 км. Мощность двигателя – 75, 90, 105, 110, 122 и 125 л.с. Тип коробки передач – АМТ, АТ, МТ. На момент исследования для продажи на автомобильных рынках находились автомобили с 2014 по 2017 г., т.е. возраст автомобиля исчисляется от 1 года до 4 лет.

Так как работа направлена на определение цены подержанного автомобиля, выходной переменной будет являться именно его цена. Изучив сайты отметим, что минимальная цена на *Skoda Rapid* приблизительно равна 400000 рублей. Максимальная цена на сайтах за *Rapid* – приблизительно 1 млн руб.

Диапазон изменения цен на автомобиль от 400000 до 1000000. Разделим данный диапазон на 6 условных отрезков по 100000, которые и будут являться выходными параметрами.

Таким образом, принимая вектор признаков из четырех параметров $G = [\text{пробег, мощность двигателя, тип коробки передач, возраст автомобиля}]$, обученная ИНС должна классифицировать автомобиль $g \in G$ путем отнесения его к одному из 6 диапазонов цен $Q = \{A, B, C, D, E, F\}$, где $A = 400000 - 500000$; $B = 500000 - 600000$; $C = 600000 - 700000$; $D = 700000 - 800000$; $E = 800000 - 900000$; $F = 900000 - 1000000$ (цена в рублях).

Архитектура и тип ИНС

В качестве ИНС используем универсальный классификатор типа «многослойный персептрон» [2, 11]. Вектор признаков состоит из четырех параметров, поэтому во входном слое 4 нейрона. Количество нейронов в выходном слое равно количеству классов, т.е. 6. В ИНС применим 2 скрытых слоя.

Известно, что можно определить число элементов в скрытом слое ИНС, при условии, что такой слой только один [12]. Это можно сделать по уравнению

$$M_s = 2 \cdot M_x + 1, \quad (1)$$

где M_x – число нейронов входного слоя, M_s – число нейронов в скрытом слое.

У нас, $M_x = 4$, поэтому вычисляем: $M_s = 2 \cdot 4 + 1 = 9$.

Если в персептроне больше одного слоя, число нейронов можно определить эмпирически [2, 11, 12]. Разместим 8 и 4 нейрона в первом и втором скрытых слоях соответственно, что суммарно больше оценки, полученной по формуле (1).

Функция гиперболического тангенса th служит функцией активации нейронов скрытых слоев персептрона Z :

$$Z(Ax_{lin}) = th(Ax_{lin}) = \frac{Sh(Ax_{lin})}{Ch(Ax_{lin})} = \frac{e^{Ax_{lin}} - e^{-Ax_{lin}}}{e^{Ax_{lin}} + e^{-Ax_{lin}}}, \quad (2)$$

где x_{lin} – сигнал линейной части искусственного нейрона, A – параметр сдвига.

Обучающие примеры

Чтобы определить необходимое число K обучающих примеров для корректного обучения ИНС методом «с учителем», применим уравнение

$$K \geq \frac{w_c}{e}, \quad (3)$$

где w_c – число настраиваемых параметров ИНС, $e \leq 1$ – приемлемое значение ошибки.

Число K берут чуть больше, как минимум на 5% от определенного в (3).

У нас: $e \approx 0,15$, $w_c \approx 90$, определяем $K \geq 600$. Полученное число увеличиваем на 5%: $K = 630$ (шт.).

Необходимое количество примеров с сайтов *auto.ru* и *drom.ru* внесем в файл, который нужен для работы *NeuroSolutions*, выполнив преобразования данных.

Обучение ИНС и анализ полученных результатов

Произведем моделирование ИНС в пакете *NeuroSolutions* в режиме «Нейроэксперт».

Количество векторов для экзаменационного и тестового наборов – 20% от общего числа. Тестовый (*Cross validation*) набор применяется во избежание «переучивания» ИНС, экзаменационный (*Testing*) – для оценки качества работы.

Задав параметры ИНС, завершаем построение сети в *NeuroSolutions* (рис. 2).

Далее, запускаем процесс обучения нейронной сети. Всего используем 1000 эпох обучения. В результате кривая обучения (*Learning curve*) принимает вид, представленный на рис. 3.

Здесь CV – среднеквадратическая ошибка на тестовом наборе, T – среднеквадратическая ошибка на обучающих данных. Обе ошибки в процессе обучения нейронной сети уменьшаются. Хорошо видно, что кривые принимают горизонтальный вид на уровне чуть менее 0,2 уже после 170 эпох обучения. Таким образом ИНС обучена и может решить представленную ей задачу со среднеквадратической точностью около 80%.

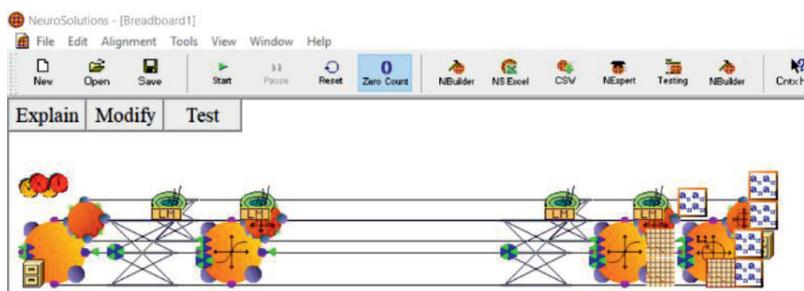


Рис. 2. Проект *NeuroSolutions* с созданной ИНС

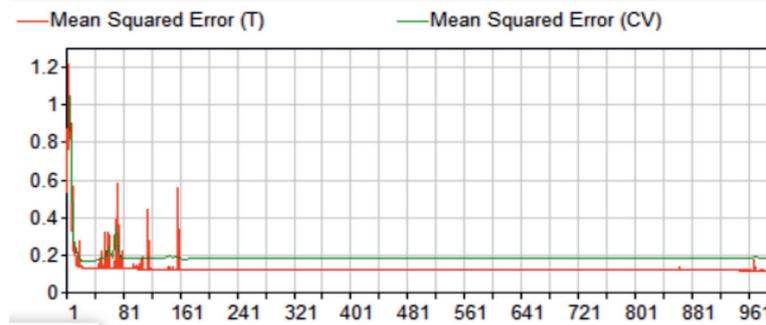


Рис. 3. Кривая обучения нейронной сети

Результаты исследования и их обсуждение

Протестируем ИНС на тестовых записях, в большинстве случаев ИНС успешно справляется с определением диапазона цены. Появление единицы в выходном векторе ИНС Q служит индикатором класса, к которому принадлежит вектор g_i .

Используемый нейропакет предоставляет средства определения общих результатов обучения ИНС на базе экзаменационного набора (*Testing*) и представления, что показано в таблице. Значения из этого набора не подавались на входы сети в процессе ее обучения и были предварительно зарезервированы в объеме 20% от общего числа примеров. Таким образом, мы можем симулировать работу ИНС в реальных условиях и оценить ее.

Итоги работы ИНС при подаче на ее входы экзаменационных данных

	A	B	C	D	E	F
A	61,77	35,29	2,94	0	0	0
B	3,73	84,47	11,18	0,62	0	0
C	0	32,5	59,17	5,83	2,5	0
D	0	6,67	37,78	44,44	8,89	2,22
E	0	0	12,5	25	56,25	6,25
F	0	0	0	0	0	100

Как видно из таблицы, наилучших результатов ИНС достигла при отнесении экземпляров автомобилей к ценовым диапазонам (классам) F и B (100% и 84,47% соответственно). Хуже всего дела обстоят с диапазоном D , где процент верных ответов составил всего 44,44%. При этом значительную долю автомобилей этой категории (37,78%) нейросеть отнесла к ближайшему (более дешевому) ценовому диапазону C .

С одной стороны, это может свидетельствовать о том, что часть автовладельцев неверно оценивает стоимость своих авто-

мобилей при определении цены продажи. С другой стороны, для оценки авто могут иметь значение другие факторы, которые мы не подавали на вход ИНС. Например, наличие скрытых дефектов или проблемы с оформлением документов.

Для улучшения результатов часть из этих параметров может быть добавлена в качестве дополнительных входов ИНС, но здесь есть опасность усложнения системы при незначительном увеличении точности классификации. Добавление слишком большого числа признаков приводит к экспоненциальному росту [13, 14] числа опорных точек, которые требуются для описания пространства (так называемое «проклятие размерности»).

Разумеется, ошибки в определении цены будут встречаться даже при работе опытных экспертов. В целом представленный результат можно считать приемлемым. На результаты ИНС можно опираться при приблизительной оценке стоимости подержанного *Skoda Rapid*.

Возможность внедрения

Применяя интегрированные в нейропакет средства, сгенерируем *DLL*-модуль, в котором находится ИНС. Это дает возможность повторно применять ее в приложениях. Для работы с *DLL* написана программа в *Visual Studio* с интерфейсом пользователя (рис. 4). Использован язык *C++*.

Приложение может работать в двух режимах: получение от ИНС результатов (*Output*) и тренировка (*Training*). В режиме «*Output*» вручную вводятся данные, нажатие кнопки «Вычислить» выводит на экран один из возможных диапазонов цен. В режиме «*Training*» возможно дообучение ИНС и запуск нового процесса обучения с обнулением межсинаптических весов сети. Это дает возможность повышать точность работы нейросети при получении дополнительных данных.

Введите характеристики и модель вашего автомобиля в следующие поля:

Марка: Модель: Год:

Пробег: Коробка передач: Мощность двигателя: л.с.

Рассчитать

Стоимость автомобиля в соответствии с ситуацией на рынке:

Стоимость:

Рис. 4. Интерфейс программного приложения

В качестве предпочтительного варианта реализации сервиса следует рассматривать его внедрение на популярных интернет-площадках по продаже автомобилей в виде *Web*-сервиса, доступного пользователям в режиме онлайн через браузер.

Заключение

В результате проведенного исследования изучена возможность применения ИНС в программных сервисах, предназначенных для оценки стоимости подержанных автомобилей.

В качестве целевой марки и модели автомобиля выбран *Skoda Rapid*. Собраны обучающие примеры с сайтов *auto.ru* и *drom.ru* в количестве 630 шт. Представлена математическая постановка задачи, как задачи классификации. В качестве ИНС использован «многослойный перцептрон» с двумя скрытыми слоями.

Произведено моделирование ИНС в *NeuroSolutions*. Проанализированы результаты работы ИНС, рассмотрены пути повышения качества ее работы и возможность внедрения. Нейросеть может использоваться для экономии времени и повышения удобства пользования.

Работа поддержана РФФИ, проекты: 18-08-00549-а, 17-07-00620-а, 19-01-00246-а, 19-07-00329-а.

Список литературы

1. Shukla N., Fricklas K. Machine learning with Tensorflow. Shelter Island: Manning Publications, 2018. 272 p.

2. Рашид Т. Создаем нейронную сеть. СПб.: ООО «Альфа-книга», 2017. 272 с.

3. Цуриков А.Н., Гуда А.Н., Карсян А.Ж. Теоретические основы интеллектуализации решения задач классификации в слабоформализуемых предметных областях // Научно-технический вестник Поволжья. 2016. № 1. С. 90–93.

4. Джоши П. Искусственный интеллект с примерами на Python. СПб.: ООО «Диалектика», 2019. 448 с.

5. Vianello F. Natural (or normal) prices. Some pointers. Political economy. Studies in the surplus approach. 1989. Vol. 5. no. 2. P. 89–105.

6. Хабибрахимов А. «Авто.ру» запустил бесплатный сервис для оценки стоимости автомобиля [Электронный ресурс]. URL: <https://vc.ru/transport/29899-avto-ru-zapustil-besplatnyy-servis-dlya-ocenki-stoimosti-avtomobilya> (дата обращения: 12.06.2019).

7. Skoda Rapid (2012) [Электронный ресурс]. URL: [https://en.wikipedia.org/wiki/%C5%A0koda_Rapid_\(2012\)](https://en.wikipedia.org/wiki/%C5%A0koda_Rapid_(2012)) (дата обращения: 14.06.2019).

8. «Skoda Auto Россия» – официальный сайт марки Шкода [Электронный ресурс]. URL: <http://www.skoda-avto.ru> (дата обращения: 15.06.2019).

9. Гамагина А.А., Цуриков А.Н. Программы для моделирования искусственных нейронных сетей: преимущества и недостатки // Актуальные проблемы и перспективы развития транспорта, промышленности и экономики России: сборник научных трудов. 2018. С. 37–39.

10. Цуриков А.Н. Применение программ-нейропакетов для решения практических задач // Инженер. 2012. № 5. С. 6.

11. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018. 480 с.

12. Tsurikov A.N., Guda A.N. Practical application of the original method for artificial neural network's training. Advances in intelligent systems and computing. 2018. Vol. 679. P. 84-93. DOI: 10.1007/978-3-319-68321-8_9.

13. Шолле Ф. Глубокое обучение на Python. СПб.: Питер, 2018. 400 с.

14. Haykin S. Neural networks: A comprehensive foundation. Upper Saddle River, New Jersey: Prentice Hall, 1999. 864 p.