

УДК 330.46:338.24

СЕМАНТИЧЕСКИЙ АНАЛИЗ ОТЗЫВОВ О ПОСТАВЩИКАХ НА ОСНОВЕ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВОЙ ТЕХНОЛОГИИ

Шимохин А.В.

ФГБОУ ВО «Омский государственный аграрный университет» имени П.А. Столыпина»,
Омск, e-mail: Schimokhin@yandex.ru

В данной статье приведены результаты работы, целью которой было исследовать возможность применения семантического анализа текста на основе нейросетевой технологии в качестве инструмента при выборе поставщика. В статье приведены успешные, практические примеры работы нейронных сетей с текстовыми данными, в области маркетинга, прогнозирования цен, работы с текстовыми документами, анализа поведения клиентов. Отмечена необходимость повышения объективности при выборе поставщика услуг за счет обработки большого объема данных. Рассмотрена возможность применения семантического анализа на основе нейронных сетей с целью поиска отрицательных и положительных отзывов с последующим формированием образа поставщика. Для обучения нейронной сети применялись корпуса твитов Ю.В. Рябцовой, символы-иконки и собранные автором отзывы о поставщиках автосервисных услуг. По этим данным разработана и обучена нейронная сеть на языке программирования Python. Эффективность данной нейронной сети была подтверждена на сообщениях-отзывах, взятых из открытых источников, что дает возможность дальнейшего применения нейросетевой технологии для отбора поставщиков, используя интернет. Такой отбор является предварительным этапом поиска поставщика, но за счет применения нейросетевой технологии становится возможным обработать не только большой массив данных, но и получить объективную выборку наиболее надежных организаций, по отзывам клиентов, за короткий период времени.

Ключевые слова: нейросетевое моделирование, нейронная сеть, поставщик, автосервис Python

SEMANTIC ANALYSIS OF SUPPLIER REVIEWS BASED ON THE USE OF NEURAL NETWORK TECHNOLOGY

Shimokhin A.V.

P.A. Stolypin Omsk State Agrarian University, Omsk, e-mail: Schimokhin@yandex.ru

This article presents the results of the work aimed at investigating the possibility of using semantic text analysis based on neural network technology as a tool when choosing a supplier. The article presents successful, practical examples of how neural networks work with text data in the field of marketing, price forecasting, working with text documents, and analyzing customer behavior. The necessity of increasing objectivity when choosing a service provider, due to the processing of a larger amount of data, is noted. The possibility of using semantic analysis based on neural networks in order to search for negative and positive reviews with the subsequent formation of the supplier's image is considered. To train the neural network, we used the corpus of Yu.V. Ryabtseva's tweets, symbols, icons, and reviews about car service providers collected by the author. Based on these data, a neural network was developed and trained in the Python programming language. The effectiveness of this neural network has been confirmed on feedback messages taken from open sources, which makes it possible to further use neural network technology to select suppliers using the Internet. Such a selection is a preliminary stage in the search for a supplier, but due to the use of neural network technology, it becomes possible to process not only a large amount of data, but also to obtain an objective sample of the most reliable organizations, according to customer reviews, in a short period of time.

Keywords: neural network modeling, neural network, supplier, Python car service

В настоящее время продолжают развиваться различные цифровые технологии, все чаще слышно о создании новых цифровых решений в различных областях человеческой деятельности. Так, например, создаются умные алгоритмы, которые могут предсказывать с большой точностью курсы валют, прогнозировать спрос и цены на товары и др. Люди различных профессий все больше освобождаются от «рутины» [1] благодаря новым цифровым технологиям [2]. Например, существуют отдельные приложения для бухгалтерского анализа, анализа спроса, прогноза цен сельскохозяйственной продукции и др. И хотя речь идет об «узком искусственном интеллекте», то есть алгоритмах, которые созданы для решения определенных задач, их эффектив-

ность не вызывает сомнений. При этом в работе [2] отмечается, что данные тенденции не приведут к исчезновению рабочих мест, а обеспечат условия для более качественной работы. Рассмотрим цифровые решения, которые, по мнению авторов [3–5], войдут в бизнес в ближайшие пять лет.

В Японии в IBM ведется разработка программы на основе нейронной сети, которая должна будет определять условия страхования клиентов. Данная программа, по заявлениям разработчиков, позволит страховым компаниям повысить продуктивность на 30% и окупится за два года [3–5].

Технология машинного обучения также применяется для борьбы с мошенничеством в различных сферах деятельности. Например, в платежной системе PayPal благо-

даря таким алгоритмам, мошеннические транзакции составляют всего 0,32%, тогда как обычно в финансовом секторе этот показатель достигает 1,32% [5].

Нейронные сети в настоящее время способны анализировать естественные языки, такие сети используются для создания чат-ботов, которые позволяют клиентам получать необходимую информацию о товарах и услугах компании. Такие алгоритмы, например, уже применяются в приемной Правительства Москвы и обрабатывают около 5% запросов [4].

В общем, цифровых решений с помощью данных технологий в областях бизнеса становится все больше.

Однако можно отметить, что различные задачи бизнеса, которые решаются с помощью нейросетевой технологии, чаще связаны с маркетингом, анализом спроса, прогнозированием цен [3]. Между тем существует ряд вопросов, связанных с организацией деятельности: выбор поставщика, поиск мест для торговли, решения об аутсорсинге и др. [6].

Так, под аутсорсингом понимается процесс передачи определенных функций и работ стороннему поставщику. Механизм передачи включает как анализ и выбор процесса для передачи на аутсорсинг, так и поиск поставщика-аутсорсера. Вообще аутсорсинг считается одним из эффективных инструментов для повышения эффективности бизнеса. Для решения задач, сопутствующих решению о аутсорсинге, менеджерам необходимо решить, по каким критериям выбирать аутсорсера или поставщика, как измерить эффективность решений, каким должен быть объем переданных работ на аутсорсинг. Для всех этих задач менеджеры со временем могут получить помощников – интеллектуальные программные продукты.

Рассмотрим рынок автосервисных услуг. По последним данным количество автомобилей, как и фирм в сферах ремонта и ТО автомобилей, растет [7]. При этом в их долю также входят услуги для корпоративных клиентов.

Как было показано в работе [6], для выбора поставщика автосервисных услуг возможно применение шкалы оценки по следующим характеристикам поставщика: время существования на рынке, отзывы о фирме-поставщике, крупные компании в клиентах поставщика. В этой же работе была предложена математическая модель

$$N_i = 0.37 \cdot R_1 + 0.23 \cdot R_2 + 0.4 \cdot R_3, \quad (1)$$

где N_i – надежность i -го поставщика услуг; R_1 – значение параметра времени существования поставщика на рынке; R_2 – зна-

чение параметра количества крупных предпочтений в клиентах данного поставщика; R_3 – значение параметра количества отзывов о фирме-поставщике.

Для параметра R_3 :

$$R_3 = \sum N_p - \sum N_n, \quad (2)$$

где R_3 – значение параметра количества отзывов о фирме-поставщике; N_p – положительный отзыв; N_n – отрицательный отзыв.

Данная модель использовалась для проверки работы нейронной сети по выбору поставщика услуг. Предполагается, что система для выполнения поставленной задачи должна уметь обрабатывать информацию о поставщиках и интерпретировать данные, как в социальных сетях, так и на сайтах, содержащих отзывы о компаниях. Актуальным становится проведение исследований возможности нейронной сети использовать семантический анализ при выполнении задачи по выбору поставщика.

Как отмечалось [8], в настоящее время ведутся исследования по разработке методов и инструментов анализа информации из массивов текстовых данных, одна из задач таких методов – извлечение ключевых понятий текстовых файлов. В случае с отзывами поставщиков большее значение приобретает эмоциональный характер сообщения, тоновая окрашенность сообщения. То есть речь идет о так называемом семантическом анализе. Название данного анализа происходит от термина семантика – дисциплина, изучающая связь слов и образов, которые передает автор текста. Семантический анализ является сложной математической задачей, которую приходится решать в процессе разработки искусственного интеллекта, результаты семантического анализа могут использоваться для анализа качества товаров или услуг, по отзывам, на сайтах, поисковиках и соцсетях.

Анализ информации о поставщиках в сети является промежуточным этапом их выбора, при котором отбираются организации, которые вызывают доверие, несут положительный образ, и именно с ними будет идти дальнейшая работа по выбору поставщика услуг. Но для повышения объективности и сокращения времени выполнения данного анализа необходимо обработать большое количество данных, чего можно достигнуть за счет применения семантического анализа на основе нейронных сетей, который определит количество отрицательных и положительных отзывов о поставщике, выберет несколько наиболее «привлекательных». После данного этапа проходит дальнейший выбор поставщика среди отобранных нейронной сетью. За счет обработ-

ки большего объема данных обеспечивается качество отбора и значительно сокращается время его выполнения. Цель исследования – изучение возможности нейронных сетей анализировать тоновую окраску отзывов о поставщике, с целью их отбора для дальнейшего выбора поставщика.

Материалы и методы исследования

С целью проверки возможности нейронных сетей анализировать тоновую окраску отзывов о поставщике была построена нейронная сеть на языке Python. Для ее обучения применялся корпус твитов Ю.В. Рябцевой [8], ее применение обосновано тем, что он содержит 114 991 размеченных коротких положительных сообщений и 111 923 размеченных коротких отрицательных сообщений. Отзывы зачастую представляют собой короткие сообщения, из которых и состоит корпус твитов, которые выражают отношение к тому или иному объекту. В работе [8] показана эффективность использования смайликов (символы-иконки, обозначающие эмоции в письменных сообщениях) для классификации текстов на положительные и отрицательные. Показано, что алгоритм более точно определяет эмоциональную окраску сообщения, если автор указал символ, обозначающий эмоции. Поэтому были составлены словари размеченных символов. Для составления словаря символов использовался ресурс Wikipedia [9]. Кроме того, в размеченный текст для обучения нейронной сети были добавлены сообщения – отзывы непосредственно об автосервисных услугах, около 300 положительных и 300 отрицательных, взятых из открытых источников [9]. Для успешного применения нейросетевой технологии из сообщений-постов были убраны знаки препинания и имена собственные. Удаление собственных имен, в данном случае названий автосервисных предприятий, обосновано тем, чтобы при обучении нейронная сеть случай-

но не «запомнила» определенное название предприятия как «плохое» или, наоборот, «хорошее» – это сделано для повышения точности оценки эмоциональной окраски поста. Допустим, отзыв плохой, но нейронная сеть оценит его как положительный из-за присутствия в нем наименования «хорошей» компании. Нейронная сеть [8] обрабатывает текстовые сообщения следующим образом: создается словарь, в котором каждому уникальному слову присваивается номер. Тогда входной слой нейронной сети будет равен количеству уникальных слов в словаре, которые активируются, когда на вход подается сообщение-отзыв. Для обучения нейронной сети использовался метод распространения обратной ошибки [9] и применялась библиотека tensorflow.

В процессе подготовки текста для обучения нейронной сети использовался модуль Stemmer (рис. 1) [8].

```
stemmer = RussianStemmer()
regex = re.compile('[^а-яА-Я ]')
stem_cache = {}

def get_stem(token):
    stem = stem_cache.get(token, None)
    if stem:
        return stem
    token = regex.sub('', token).lower()
    stem = stemmer.stem(token)
    stem_cache[token] = stem
    return stem
```

Рис. 1. Реализация модуля Stemmer

Данный модуль позволяет работать с кириллическими символами и отсекает от слов окончания и суффиксы. В результате из собранной обучающей выборки было получено 92 200 уникальных слов. Из них для словаря нейронной сети было отобрано 5 100 слов по повторяемости в сообщениях-отзывах. Алгоритм построения нейронной сети показан на рис. 2.

```
def build_model(learning_rate=0.1):
    tf.reset_default_graph()

    net = tflearn.input_data([None, VOCAB_SIZE])
    net = tflearn.fully_connected(net, 125, activation='relu')
    net = tflearn.fully_connected(net, 25, activation='relu')
    net = tflearn.fully_connected(net, 2, activation='softmax')
    regression = tflearn.regression(
        net,
        optimizer='sgd',
        learning_rate=learning_rate,
        loss='categorical_crossentropy')

    model = tflearn.DNN(net)
    return model
```

Рис. 2. Алгоритм построения нейронной сети

Результаты исследования и их обсуждение

Для нейронов промежуточных слоев, по рекомендации [8], была выбрана функция активации Relu. Ее рекомендуют использовать при большом количестве нейронов. Проверка точности распознавания на тестовой выборке показала точность нейронной сети 73%.

Затем для проверки были взяты некоторые сообщения из выборки, и нейронная сеть в целом правильно их охарактеризовала (рис. 3–4).

В последнем примере, несмотря на то, что в отзыве было около 15% незнакомых слов, нейронная сеть правильно охарактеризовала его как отрицательный. На рис. 5 показан ввод новых отзывов и результат их оценивания нейронной сетью.

```
test_tweet_number(222)
Original tweet: Ребята знают своё дело. Все на высшем уровне. А главное ВОВРЕМЯ! Всем советую.
P(positive) = 0.94681. Result: Positive
```

Рис. 3. Позитивный отзыв (95%)

```
Unknown token: вариатор
Unknown token: рекламе
Unknown token: идет
Unknown token: обвинили
Unknown token: чуть
Unknown token: выгонять
Unknown token: хамы
Unknown token: вруны
Unknown token: лгуны
Unknown token: некто
Unknown token: кому
Unknown token: водителям
Unknown token: разведут
Original tweet: приехал на данное СТО и это за два года третий раз стали менять масло в АКПП вариатор у них все есть фильтра и масла и по рекламе идет акция работа бесплатно деньги все равно взяли за работу в ходе работы сделал замечание мастеру что кпп не помыли и стали менять масло и фильтра во всем обвинили клиента точнее меня раз я не помыл акпп я объяснил что у Вас есть мойка мне объяснили что мойка не для мойки и стали чуть не силой выгонять из бокса мое мнение хамы вруны и лгуны так плохо к моей ласточки за 15 лет некто не относился в боксах грязь не кому не советую это СТО особенно водителям женщинам разведут на разные услуги
P(positive) = 0.03463. Result: Negative
```

Рис. 4. Отрицательный отзыв на 97%

```
In [97]: tweets_for_testing = [
        " В автокомплекс на улице Химиков обратился за покупкой запчастей. Приобрёл здесь тяги рулевые. Всё понравилось.Сотрудники
        "После обслуживания авто в данном СТО, а именно замена масла в ДВС, спустя 3 месяца заметил течь масла с днище машины.
        ]
for tweet in tweets_for_testing:
    test_tweet(tweet)
    print("-----")

Unknown token: Химиков
Unknown token: покупкой
Unknown token: Приобрёл
Unknown token: товару
Unknown token: Выдали
Unknown token: ассортимент
Unknown token: чуть
Original tweet: В автокомплекс на улице Химиков обратился за покупкой запчастей. Приобрёл здесь тяги рулевые. Всё понравилось
.Сотрудники всё рассказали, объяснили по товару, помогли сделать выбор. Выдали чек, гарантии по чеку. В этом автокомплексе х
ороший ассортимент. Всё есть в наличии. По цене хочется всегда чуть дешевле.
P(positive) = 0.92827. Result: Positive
-----
Unknown token: именно
Unknown token: течь
Unknown token: днище
Unknown token: закрутили
Unknown token: масляный
Unknown token: фактор
Unknown token: считаю
Unknown token: недопустимым
Unknown token: похоже
Unknown token: отмывание
Unknown token: повторном
Unknown token: коллегам
Unknown token: будут
Unknown token: отрицательными
Original tweet: После обслуживания авто в данном СТО, а именно замена масла в ДВС, спустя 3 месяца заметил течь масла с дни
ще машины. Оказалось не закрутили масляный фильтр. Данный фактор считаю недопустимым, т. к. это просто похоже на отмывание денег
при повторном обращении к ним.В данный сервис более обращаться не буду, рекомендации друзьям, коллегам и знакомым так же будут к
райне отрицательными.
P(positive) = 0.04701. Result: Negative
-----
```

Рис. 5. Оценка отзывов из открытых источников

Как видим, несмотря на существенное количество незнакомых слов, нейронная сеть смогла правильно охарактеризовать тональную окраску сообщений.

Заключение

Таким образом, изучена возможность нейронных сетей применяться в качестве инструмента для семантического анализа отзывов о поставщиках автосервисных услуг. Для обучения созданной нейронной сети использовались корпуса твитов [9], расширенные отзывами о фирмах-поставщиках автосервисных услуг, собранных автором. Функция активации для скрытых слоев использовалась Relu, продемонстрирована возможность нейронной сети работать с большим массивом данных об отзывах компаний предоставляющих услуги автосервиса, в том числе для корпоративных клиентов, что в целом необходимо для получения характеристики поставщика. Анализ информации о поставщике в интернете становится одним из ключевых факторов отбора поставщика, но для объективности и уменьшения времени выполнения такого анализа необходимо обработать большое количество данных, в статье показано, что для этого возможно применить семантический анализ на основе нейронных сетей.

Список литературы

1. Шимохин А.В., Хаирова С.М., Хаиров Б.Г., Алгазин Д.Н. Тенденция изменений трудовых функций в связи с развитием цифровых технологий // Вестник Алтайской академии экономики и права. 2020. № 11–3. С. 573–578.
2. Русанов И.А., Четвериков А.В. Современное состояние и перспективы развития искусственного интеллекта // Научные технологии и инновации: Международная научно-практическая конференция (XXII научные чтения). 2016. С. 56–61.
3. Екимова К.В., Лукьянов С.А., Смирнова Е.Н. Цифровая экономика и искусственный интеллект: новые вызовы современной мировой экономики: монография. М.: Издательский дом ГУУ, 2019. 180 с.
4. Шотыло Д.М., Крайнова В.Е., Скурадин А.В. Тенденции развития искусственных нейронных сетей в цифровой экономике // ЭКОНОМИНФО. 2018. № 4. С. 65–69.
5. Крипто-климат. Нейросети: как искусственный интеллект помогает в жизни и бизнесе [Электронный ресурс]. URL: <https://dioklimat.ru/servisy-i-ploshchadki/primenenie-iskusstvennogo-intellekta-2.html> (дата обращения: 11.04.2021).
6. Шимохин А.В. Применение нейросетевого моделирования для принятия решения о передаче бизнес-процесса на аутсорсинг // Фундаментальные исследования. 2019. № 5. С. 160–166.
7. Лузина М. Объем российского рынка автосервиса по ТО и ремонту // Автостат. Аналитическое агентство [Электронный ресурс]. URL: <https://yandex.ru/turbo/autostat.ru/s/infographics/41081/> (дата обращения: 20.03.2021).
8. Пройдаков Э.М. Современное состояние искусственного интеллекта // Научные исследования. 2018. С. 129–153.
9. Рубцова Ю.В. Метод построения и анализа корпуса коротких текстов для задачи классификации отзывов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XV Всероссийской научной конференции RCDL'2013. 2013. С. 269–275.