



УДК 004.048

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ОНЛАЙН-РЕПУТАЦИЕЙ НА ОСНОВЕ ЛЕКСИЧЕСКОГО АНАЛИЗА ОТКРЫТЫХ ДАННЫХ ТОРГОВЫХ ПЛОЩАДОК

*М.А Косоногова, marinakosonogovasc@gmail.com, А.В. Голиков,
golikov-it@yandex.ru, Р.У. Стативко, stativko1@mail.ru,*

Белгородский государственный технологический университет им. В.Г. Шухова, г. Белгород

В данной статье рассматривается разработка системы, направленной на формирование онлайн-репутации торговых площадок. Работа системы построена на продуцировании отзывов, позволяющих пользователю получить целостное представление о товаре на основе интеграции опубликованных мнений о каждом из его параметров. Достижение результата осуществляется за счет внедрения технологии машинного обучения.

Ключевые слова: машинное обучение, онлайн-торговля, отзывы, автоматизация.

THE DEVELOPMENT OF INTELLIGENT SYSTEM MANAGEMENT OF ONLINE REPUTATION-BASED LEXICAL ANALYSIS OF OPEN DATA MARKETPLACES

M.A. Kosonogova, A.V. Golikov, R.U. Stativko

Shukhov Belgorod State Technological University, Belgorod

This article discusses the development of a system aimed at the formation of online reputation of trading platforms. The work of the system is based on producing reviews that allow the user to get a holistic view of the product based on the integration of published opinions about each of its parameters. The result is achieved through the introduction of machine learning technology.

Keywords: machine learning, online trading, reviews, automation.

Интернет-торговля на сегодняшний день переживает период активного развития. По данным Ассоциации компаний интернет-торговли, в минувшем году объем онлайн-рынка вырос на 21%. Электронная торговля уже составляет 36% всей цифровой экономики России, и этот показатель, судя существующим тенденциям, в будущем будет активно увеличиваться [1]. Под электронной коммерцией, вслед за А.В. Юрасовым, мы будем понимать «любой вид сделок, при которых взаимодействие сторон осуществляется электронным способом вместо физического обмена или непосредственного физического контакта» [2, с. 12].

Развитие популярности интернет-торговли приводит к активному росту количества онлайн-магазинов и торговых площадок. В свою очередь, это становится причиной появления конкуренции между ритейлерами. Стремясь завоевать расположение потенциальных покупателей, они используют самые разные методы: проведение рекламных компаний, выстраивание гибкой ценовой политики, расширение ассортимента и т.п. Одним из продуктивных приемов также способно стать совершение технологии работы торговой площадки. Добавление новых элементов, способных сделать процесс выбора и покупки товара более простым и удобным, позволяет существенно увеличить лояльность покупателей и, как следствие, повысить продажи.

Целью проекта стала разработка системы, позволяющей автоматически генерировать возможные отзывы на основе технических характеристик товара. Технология машинного обучения обуславливает возможность постоянного совершенствования



системы и повышения точности предлагаемых пользователю оценок. Система позволит определять функциональность и продуктивность работы технически сложного товара на основе его спецификации. Оценка каждого параметра производится на базе отзывов, оставленных пользователем Интернета на различных ресурсах. В результате готовый продукт будет полезен потенциальным покупателям, выбирающим оптимальный для себя товар. Система может использоваться как автономно, так и в пределах крупного онлайн-магазина или маркетплейса.

Актуальность разрабатываемого проекта обусловлена уже упоминавшимся ростом популярности интернет-торговли. В поисках новых покупателей торговым площадкам необходимо постоянное совершенствование используемых технологических решений, что невозможно без внедрения новых компонентов. Наряду с этим нельзя забывать и об обилии товаров, предлагаемых онлайн-магазинами. Стремясь выбрать оптимальный вариант, пользователи не только анализируют спецификации интересующих их позиций, но и стремятся познакомиться с мнениями тех, кто уже воспользовался товаром.

Значимость отзывов как средства регулирования онлайн-репутации сложно переоценить. Неслучайно в пределах каждой торговой площадки можно обнаружить модуль, позволяющий ознакомиться с информацией, оставленной пользователями. Крупные маркетплейсы формируют рейтинги товаров и интернет-магазинов, что существенно влияет на решение потенциального покупателя о выборе конкретной товарной позиции и онлайн-магазина. Неслучайно недобросовестные продавцы нередко пользуются ложными отзывами, искусственно улучшающими репутацию как самих торговых площадок, так и отдельных товаров.

Системы, генерирующие недостоверные отзывы, на сегодняшний день не редкость. Обращает на себя внимание отсутствие ресурса, работающего с реальной информацией и генерирующего синтетический отзыв на основе собранных пользовательских мнений и технических характеристик продукта. На устранение данной лакуны и направлен рассматриваемый проект. Впервые, благодаря внедрению технологии машинного обучения, удастся создать систему, самостоятельно генерирующую информативные и грамматически верные отзывы на базе опубликованных пользовательских мнений.

Разрабатываемый ресурс избавляет от необходимости самостоятельного мониторинга сайтов, содержащих отзывы. Система собирает все пользовательские отклики, извлекая из них ключевую информацию. Генерируется новый синтетический отзыв, позволяющий ознакомиться с максимально объективными сведениями о товаре.

Работа по созданию предлагаемой системы проводится в несколько этапов:

1. Анализ рынка интернет-магазинов с целью выявления наиболее приоритетных категорий товаров для первичной реализации проекта.
2. Разработка методов и алгоритмов сбора информации о технических характеристиках товара и опубликованных пользовательских отзывов.
3. Разработка структуры базы данных для хранения в унифицированном виде информации по п.2 из разных источников.
4. Проектирование и обучение нейронной сети для анализа и обработки полученных данных.



5. Создание минимально жизнеспособного прототипа системы.

В настоящий момент и в России, и в мире отсутствуют прямые аналоги предлагаемого продукта. При этом использование технологии машинного обучения при работе с отзывами пользователей не является новым. Так, компания IBM разместила открытый код, помогающий отличать реальные отзывы от мошеннических при помощи *deep learning*. Настоящее решение имеет узко ориентированную сферу применения и направлено лишь на ресторанный бизнес. Данные о реализации указанного выше кода в конкретных программных продуктах, а также об их успешности, отсутствуют.

В России ближайшим аналогом предлагаемой системы можно считать Яндекс.Маркет. Опираясь на содержание отзывов и информацию, размещаемую в поле «Достоинства», маркетплейс вычленяет ключевых характеристики, показавшиеся пользователю наиболее полезными. Путем статистических подсчетов, таким образом, вычленяются параметры товара, пользующиеся наибольшей популярностью. Два из них включаются в подзаголовок товара, позволяя потенциальному покупателю увидеть конкретное преимущество той или иной модели. Минусом подобной системы становится недостаточно полная информация о совокупной пользовательской оценке: отображаются лишь самые популярные характеристики, тогда как остальные параметры остаются вне поля зрения. В то же время предлагаемая система подвергает анализу все спецификации и позволяет отобразить развернутые совокупные отзывы как по всему товару, так и по отдельным его параметрам.

Синтетический отзыв, генерируемый системой, формируется в соответствии с четкими критериями:

1. Соответствие нормам русского литературного языка: использование нейтральной общеупотребительной лексики, следование законам грамматики (согласование слов в роде, числе и падеже, построение верных предложений и т.п.), смысловая связность. Достигается средствами *rumorphy2*.

2. Основа отзыва – набор тезисов, сформулированных на основе технических характеристик товара.

3. Отзывы генерируются **ИСКЛЮЧИТЕЛЬНО** на основе реальных пользовательских оценок, находящихся в свободном доступе в интернете.

4. Обязательная черта совокупного отзыва – информативность. Генерируемый текст включает в себя только те сведения, которые могут так или иначе повлиять на выбор пользователя и быть для него полезными. Различные эмоциональные отклики в процессе обработки исходных отзывов переосмысливаются и используются как элемент оценки.

Для реализации технологии машинного обучения был выбран метод обучения с учителем. Это требует решения целого ряда задач: классификации, регрессии, рекомендации, заполнении пропусков и т.д. Обучение производится на основе предварительно сформированной базы данных. Предполагается, что работа будет производиться с большим массивом данных – Big Data, что представляет собой «целую серию инструментов и методов для обработки огромных структурированных и неструктурированных объемов данных разнообразных форматов и находящихся в различных местах» [3. с. 87].



Для начальной разработки системы и получения первых результатов была выбрана такая категория товаров, как смартфоны. Это обусловлено сразу несколькими факторами, в частности:

- популярность данной категории в рамках интернет-торговли;
- широкий ассортимент;
- обилие пользовательских отзывов;
- оптимальный набор технических характеристик, влияющих на решение потенциального покупателя о покупке.

Основные спецификации были сгруппированы в несколько кластеров:

- общие – группа основных товарно-технических характеристик (производитель, цена, операционная система, количество SIM-карт, год выпуска, объем встроенной памяти);
- рейтинг на бенчмарке (использовались показатели наиболее популярного бенчмарка AnTuTu);
- экран – характеристики, влияющие на свойства экрана (диагональ, разрешение, PPI, соотношение сторон);
- камера – спецификации, от которых зависит качество снимков (количество мегапикселей в тыловой и фронтальной камерах, разрешение видео, автофокус). В данном кластере рассматривались только те характеристики, которые присутствуют в пользовательских отзывах (более сложные спецификации вроде особенностей диафрагмы или фокуса были опущены);
- аккумулятор – характеристики, влияющие на длительность работы смартфона в режиме ожидания и разговора (емкость, тип);
- корпус – характеристики, оказывающие влияние на внешние особенности устройства (тип, материал, вес).

Обработка позиций по приведенной схеме позволяет сформировать следующую базу данных [Таблица 1].

Результатом обработки приведенных выше характеристик становится генерация следующих отзывов [Таблица 2].

Очевидно, что сгенерированные отзывы позволяют не просто получить совокупную оценку каждой модели, но и помогают детально узнать ее достоинства и недостатки с опорой на ключевые спецификации. Это дает возможность принять взвешенное решение о покупке и избавиться от необходимости мониторинга многочисленных отзывов в интернете.

Таким образом, предлагаемый проект имеет большое значение для Интернет-торговли, т.к. позволяет упростить процесс приобретения товара, а также дает продавцам возможность отбора позиций, потенциально привлекательных для клиента. Это позволяет повысить уровень продаж и улучшить репутацию торговой площадки.



Таблица 2

Синтетические отзывы

1			Синтетический отзыв
2	Модель	Производитель	
3	LeEco Le Pro 3 X651	LeEco	Камера слабовата; Собран качественно; Мощный аппарат 821dragon; громкие динамики; Работает очень быстро; нет возможности увеличить память;
4	Xiaomi Mi Note 3 6/128Gb	Xiaomi	Экран отличный; Хорошая детализация; Все летает; большая батарейка; Скользкий без чехла; Нет миниджека; Отличное железо; Очень шустрый

Список используемой литературы

1. Рынок интернет-торговли в 2017 году, данные АКИТ [Электронный ресурс] // E-pepper. URL: <https://e-pepper.ru/news/rynok-internet-torgovli-v-2017-godu-dannye-akit.html> (дата обращения: 25.12.2018).
2. Юрасов А.В. Электронная коммерция: Учеб. пособие. – М.: Дело, 2003. – 480 с.
3. Царькова Н.С., Смолянов А.С. Bigdata. Развитие, анализ и технологии // Актуальные проблемы гуманитарных и естественных наук. – 2016. – №7-1. – С. 86-95.