

**К.т.н. Рудниченко Н.Д., д.т.н. Вычужанин В.В., к.т.н. Шibaева Н.О.,
Шibaев Д.С.,**

**ОЦЕНКА ТОНАЛЬНОСТИ ОТЗЫВОВ ПОЛЬЗОВАТЕЛЕЙ НА
ОСНОВЕ ИСПОЛЬЗОВАНИЯ МАШИННОГО ОБУЧЕНИЯ**

**Ph.D. Rudnichenko N.D., Dr.Sc. Vychuzhanin V.V., Ph.D. Shibaeva N.O.,
Shibaev D.S.**

**ESTIMATION OF THE USER REVIEWS TONALITY USING
MACHINE LEARNING**

В настоящее время в информационной среде Internet наблюдается рост объемов разнородных данных, что связано с развитием и распространением социальных сетей, интернет-магазинов, тематических блогов и информационных веб-систем, что существенно отражается на активности развития разных направлений электронной коммерции и торговли различной электронной продукцией (EG) в частности [1]. В связи с регулярным появлением и активным развитием новых коммерческих и информационных ресурсов современные потребители виртуальных и физических товаров и услуг все чаще испытывают трудности в выборе компаний, организаций, производителей конкретных моделей технических гаджетов и средств. Это создает необходимость получения дополнительной информации о реальных функциональных возможностях и особенностях эксплуатации EG со стороны других пользователей и опытных экспертов [2]. Дополнительные сложности вносит необходимость проведения фильтрации и анализа маркетинговых мероприятий конкурирующих компаний-производителей для выявления наиболее подходящих товаров и услуг для конкретных нужд пользователя, для чего требуется применение различных методов интеллектуального анализа данных [3]. С целью получения конкурентных преимуществ и лучшего понимания запросов клиентов организации вендоры также нуждаются в оперативном получении максимально достоверных и актуальных данных на базе проведения анализа мнений пользователей.

Целью данной работы является исследование специфики применения искусственных нейронных сетей (ИНС) в качестве метода машинного

**Матеріали VIII Міжнародної науково-практичної конференції
«Інформаційні управляючі системи та технології»
23 - 25 вересня 2019, Одеса**

обучения для автоматизации процесса оценки тональности пользовательских отзывов на электронные товары.

Технически, задача оценки тональности пользовательских отзывов, представляющая собой разновидность категории задач обработки естественного языка, сводится к решению задачи многоклассовой классификации, для чего эффективно использование различных моделей ИНС [4,5]. В рамках предлагаемого подхода точность решения задачи классификации предлагается увеличить путем проведения дополнительных процедур по текстовой обработке:

1. Приведение всех встречающихся в тексте символов к нижнему регистру с целью уменьшения числа уникальных терминов в словаре.

2. Исключение из текста не буквенных символов. Подобная процедура значительно понижает число уникальных терминов в словаре, в тех случаях, когда для текста характерно обилие пунктуации, не несущей принципиальной смысловой нагрузки.

3. Исключение повторяющихся символов. Это позволяет заменить имеющиеся в тексте последовательности одинаковых символов для снижения размерности словаря.

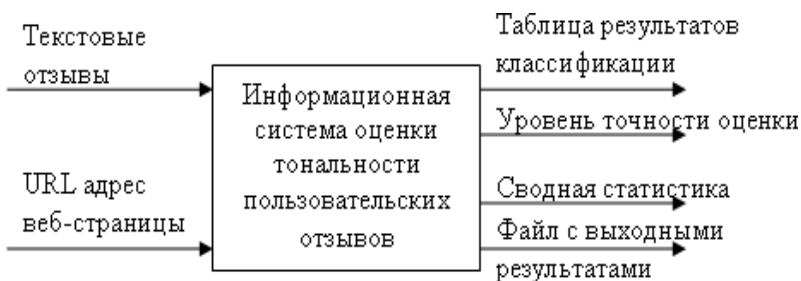


Рис. 1. Обобщенная схема входных и выходных данных ИС

4. Выделение основы слова из набора входных текстовых данных.

Перечисленные действия необходимо выполнять перед процессом классификации текста, чтобы увеличить скорость и уменьшить итерационную и логическую сложность обработки данных. Разработанное программное обеспечение в виде информационной системы (ИС) реализации оценки тональности отзывов позволяет проводить автоматический анализ входных текстов, предоставляя на выходе тип проанализированного текста: положительный, негативный или нейтральный. На вход ИС поступают текстовые данные пользовательских комментариев и отзывов, в результате обработки формируется таблица классов текста, рассчитывается уровень точности оценки (погрешность

**Матеріали VIII Міжнародної науково-практичної конференції
«Інформаційні управляючі системи та технології»
23 - 25 вересня 2019, Одеса**

классификации моделью ИНС), форма сводной статистики и файл с выходными результатами классификации в формате *.xls (рис. 1).

Выводы. Разработанное программное обеспечение поддержки процессов оценки пользовательских отзывов является кроссплатформенным и обеспечивает достаточно высокую точность классификации, более 92%, что свидетельствует о достоверности решения задачи.

Литература

1. Рудниченко Н.Д. Разработка проекта модуля конвертации видео-файлов системы интеллектуального анализа больших объемов данных / Н.Д. Рудниченко, В.В. Вычужанин, Н.О. Шибаева, Д.С. Шибаев // Сучасні інформаційні технології, засоби автоматизації та електропривод. Інтелектуальний аналіз даних та знань (Data Mining) 2018. – Режим доступа: <http://dspace.dgma.donetsk.ua/handle/DSEA/380>. – Дата доступу: 15.01.2019.

2. Рудніченко М.Д. Проектування модуля обробки мультимедійних файлів для аналізу і вилучення знань / М.Д. Рудніченко, Н.О. Шибаєва // Naukowy i innowacyjny potencjał prezentacji: kolekcja prac naukowych «ЛОГОС» z materiałami Międzynarodowej naukowo-praktycznej konferencji, Opole. Równie: «Volynsky Oberegi» Publishing House, 2018. – Т. 7. – С.90 – 92.

3. Щербина А.Д. Порівняльний аналіз існуючих напрямів у інтелектуальному аналізі даних / А.Д. Щербина, Д.С. Шибаєв, М.Д. Рудніченко, Н.О. Шибаєва // Project, Program, Portfolio Management. The Third International Scientific-practical Conference, Odesa, ONPU, 07–08 Dec 2018. – С.88 – 90.

4. Васильев В.Г. Классификация отзывов пользователей с использованием фрагментных правил / В.Г. Васильев, М.В. Худякова, С. Давыдов // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая–3 июня 2012 г.), 2012. – Т1. – С. 66 – 78.

5. Гаршина В. В. Разработка системы анализа тональности текстовой информации / В. В. Гаршина, К. С. Калабухов, В. А. Степанцов, С. В. Смотров // Вестник ВГУ, серия: системный анализ и информационные технологии, 2017. – № 3. – С.185 – 194.