

**Маньков Александр Сергеевич,**

студент,  
Кафедра анализа систем и принятия решений,  
Институт экономики и управления,  
ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н.Ельцина»,  
г. Екатеринбург, Российская Федерация

**Детков Александр Александрович,**

Доцент, Кандидат экономических наук,  
Кафедра анализа систем и принятия решений,  
Институт экономики и управления,  
ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н.Ельцина»,  
г. Екатеринбург, Российская Федерация

**АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТА. ОБЗОРНАЯ СТАТЬЯ***Аннотация:*

За последние годы количество информации в интернете значительно увеличилось и с каждым днем информации становится все больше, причем ежедневное обновление достаточно уникально, так как в мире происходит огромное количество разных действий и событий. Средства массовой информации трактуют все происходящие события в определенных стилях для разных групп читателей. Анализ этого текста позволяет понять, о чем говорил автор и как он к этому относится.

*Ключевые слова:*

Анализ тональности текста, анализ текста, машинное обучение, нейронные сети, анализ, автоматическая обработка естественного языка, извлечение мнений, субъективность, оценка.

Благодаря анализу тональности текстов, можно выяснить его эмоциональную составляющую. Это может пригодиться как для выбора фильма для просмотра, так и для крупных и значимых проектов в компании. Одним из таких может быть умный автоответчик или бот, например, в банке, который ответит на любой вопрос.

При правильном формировании условий, бот сможет понять не только суть текста, но и его эмоциональность, тем самым представить полный смысл передаваемого. На основании этого, дать более развернутый и понятный человеку ответ.

Анализ тональности текста могут трактовать по-разному. Рассмотрим два определения.

Анализ тональности текста (англ. Sentiment analysis) – это задача компьютерной лингвистики, заключающаяся в определении эмоциональной окраски (тональности) текста и, в частности, в выявлении эмоциональной оценки авторов по отношению к объектам, описываемым в тексте [1].

Компания Амазон приводит следующее определение. Анализ тональности – это процесс анализа цифрового текста для определения того, является ли эмоциональный тон сообщения положительным, отрицательным или нейтральным [2].

Оба определения говорят про анализ цифрового текста и выявлении эмоциональности, то есть оценки текста по его смыслу.

Исследование текста на сегодняшний день является обязательной необходимостью для любой коммерческой компании. Письма на электронной почте, социальные сети, комментарии, отзывы и другие текстовые данные представляют собой огромную базу полезной информации, которую необходимо анализировать.

Инструменты анализа тональности позволяют обрабатывать тексты и автоматически определять отношения автора к теме. Дальнейшее использование результатов анализа может быть реализовано в развитии отношений с клиентами или для повышения доверия к бренду компании.

Автоматическая обработка текста позволяет избежать предвзятого отношения со стороны исследователя и получить объективную информацию при проведении работ по анализу текстовых данных.

Если рассматривать анализ тональности текста как определение эмоциональной составляющей, можно обратить внимание на социальные сети. В этой области анализ автоматически определит эмоциональную оценку автора текста [3]. Данное направление анализа может быть использовано для проведения маркетинговых исследований, благодаря которым будет выделена полноценная картина лояльности клиентов каких-либо сервисов или найден процент недовольных посетителей сайта.

В данном случае все тексты будут иметь характерный разговорный стиль. Использование сленга, упрощенные слова, различные опечатки и хаотично составленные предложения будут значительно затруднять анализ текста [4].

Как и с обычными текстами, анализ текстов социальных сетей для определения эмоциональности будет задействована бинарная классификация, то есть разделение на позитивный и негативный результат. Возможно использование третьего классификации, нейтрального класса [5].

В исследовании выделяют три основных метода анализа тональности текста. Они схожи с описанными в других источниках.

Первый из них – это подходы, основанные на правилах, которые определяются вручную на основе классификации текста. За основу берутся эмоциональные ключевые слова и их совместное использование в тексте с другими ключевыми словами. Создание таких правил достаточно трудоемкий процесс, а производительность не столь высока.

Для русского языка выбор данного метода особенно сложен, так как в отличие от английского языка, для русского языка не так много ресурсов в области анализа тональности текста.

Лексиконы настроений RuSentiLex [7] и LINIS Crowd [8] являются самыми крупными лексиконами для русского языка. Однако они содержат информацию только об уровне настроений от положительного до отрицательного без каких-либо характеристик, специфичных для эмоций.

Второй тип – это подходы, основанные на машинном обучении, которые автоматически извлекают признаки из текста с последующим применением алгоритмов классификации машинного обучения, таких как: наивный байесовский классификатор, дерево решений, логистическая регрессия и машина опорных векторов. В одной из статей также подтверждается более частое использование этих методов, нежели традиционных подходов в анализе тональности текста [8].

Этот факт также был подтвержден хронологией конкурса SemEval, где ведущие решения успешно использовали сверточные нейронные сети (CNNs) и рекуррентные нейронные сети (RNNs) [9].

Третий – это гибридные подходы, которые сочетают в себе подходы, основанные как на правилах, так и на машинном обучении. Хорошим примером будет реализованная и описанная платформа анализа тональности текста, которая объединила лингвистические правила и модули CNN и LSTM для классификации настроений [10].

Сочетание эффективных действий в подходах, основанных на правилах и методов, основанных на машинном обучении, позволяют получать более точные результаты. С другой стороны, в объединенном варианте использования накладываются минусы, которые присутствуют в первом и втором подходах.

Одним из методов анализа текста, который относится к машинному обучению, является использование сверточной нейронной сети (Convolutional Neural Networks). Для построения нейронной сети можно использовать библиотеку Keras, которая является высокоуровневой надстройкой над TensorFlow [15].

Для работы CNN необходима матрица, которая может быть получена после преобразования слов из текста в вектора. Затем вектора необходимо выстроить в ряд и получится необходимая матрица. Для осуществления описанных действий необходимо использовать метод embedding, он позволяет сопоставить точки в каком-то многомерном пространстве объекту, в нашем случае – слову [16]. На рисунке 2 представлена информация как применяются свертки к тексту.

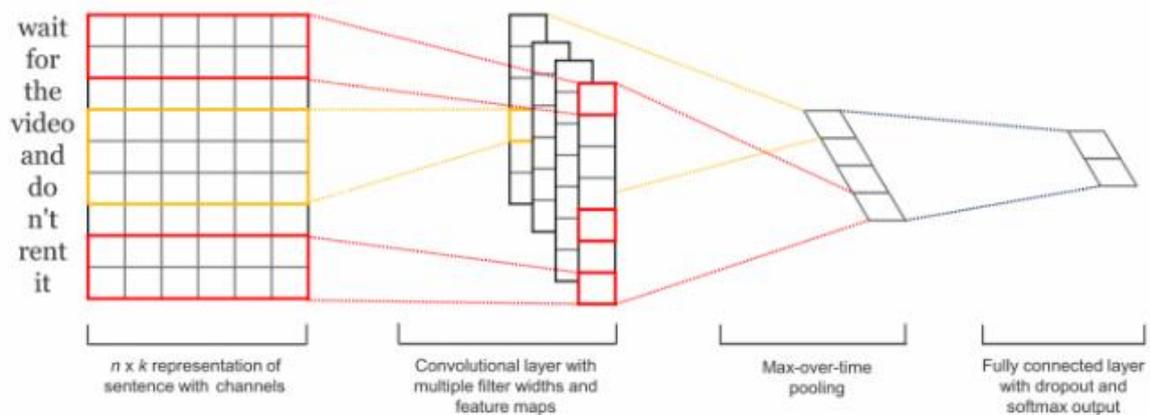


Рисунок 2 – Применение сверток к тексту.

Данный метод анализа текста описан в одной из статей, где автор рассматривает другие работы по анализу текста и выделяет проблему классификации [17]. Для решения необходимо представление объекта в виде векторов признаков. Тексты имеют неоднородную структуру, поэтому необходима предварительная обработка для построения векторов и использования метода классификации.

Результатом предварительной обработки является чистая нормализованная последовательность токенов, которые можно сопоставить с векторами признаков.

В настоящее время используются как оба подхода по отдельности, так и третий, смешанный подход. Все больше в задачах анализа текста набирают популярность нейронные сети.

Анализ тональности текста представляет собой сложную область изучения с множеством различных препятствий. Главной проблемой является обработка естественного языка компьютером, в особенности, анализ русского языка.

Анализ текста для обычного человека может показаться не столь важным занятием. Рядовому обывателю интернета и так понятно, какая эмоциональность у текста, который он читает. Однако практическое применение может быть найдено в реализации фильтра информационного потока.

Фильтрация по критерию отбора только необходимой информации позволит находить нужные тексты и статьи, или же сортировать данные на позитивные, негативные или нейтральные для дальнейшей работы с ними. Это позволит избавиться, например, от негативных новостей, которые плохо воздействуют на многих читателей.

Для бизнеса такой анализ будет действительно полезным и позволит оптимизировать расходы на обработку информации. Различные боты в мессенджерах, социальных сетях и колл-центрах обрабатывают и систематизируют информацию.

Не менее интересным направлением будет реализация анализа тональности текста на различных онлайн-площадках и конференциях, где есть открытый чат для всех участников и зрителей. Обрабатывать информацию вручную будет нецелесообразно, автоматическая система обойдется дешевле и справится с работой эффективнее.

### СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. ifmo: Классификация текстов и анализ тональности, [Электронный ресурс] [https://neerc.ifmo.ru/wiki/index.php?title=Классификация текстов и анализ тональности](https://neerc.ifmo.ru/wiki/index.php?title=Классификация_текстов_и_анализ_тональности)
2. amazon web services: Что такое анализ эмоциональной окраски, [Электронный ресурс] <https://aws.amazon.com/ru/what-is/sentiment-analysis/>
3. Pang B., Lee L. Opinion mining and sentiment analysis // Foundations and Trends in Information Retrieval. 2018. Vol. 2. P. 1-135
4. Boudad N., Faizi R., Oulad Haj Thami R., Chiheb R. Sentiment analysis in Arabic: A review of the literature // Ain Shams Engineering Journal. 2018. Vol. 9, № 4. P. 2479-2490
5. Tartir, pp. & Abdul-Nabi, I. (2017) Semantic sentiment analysis in arabic social media. Arabic Natural Language Processing: Models, Systems and Applications. 29 (2). pp. 229-233

#### **Mankov Alexander S.,**

Master student,

Department of Systems Analysis and Decision Making,

Institute of Economics and Management,

Federal State Educational Institution of Higher Professional Education «Ural Federal University named

after the First President of Russia B.N. Yeltsin»,

Yekaterinburg, Russian Federation

#### **Detkov Alexander A.,**

Associate Professor, Candidate of Economic Sciences,

Department of Systems Analysis and Decision Making,

Institute of Economics and Management,

Federal State Educational Institution of Higher Professional Education «Ural Federal University named

after the First President of Russia B.N. Yeltsin»,

Yekaterinburg, Russian Federation

### SENTIMENT ANALYSIS OF TEXT. REVIEW ARTICLE

#### *Abstract:*

In recent years, the amount of information on the Internet has increased significantly and every day there is more and more information, and the daily update is quite unique, since a huge number of different actions and events take place in the world. The media interprets all the events in certain styles for different groups of readers. An analysis of this text allows you to understand what the author was talking about and how he relates to it.

#### *Keywords:*

Sentiment analysis of text, sentiment analysis, machine learning, neural networks, analysis, automatic natural language processing, opinion extraction, subjectivity, evaluation.