

УДК 004.9+81

АНАЛИЗ МЕТОДОВ ВЫЯВЛЕНИЯ ЭМОЦИОНАЛЬНОГО КОНТЕКСТА В ТЕКСТАХ

Павлюченко К.А., магистрант

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Парамонов А.И. – канд. техн. наук, доцент, зав. каф. ИСиТ

Аннотация. В работе исследуется проблема выявления эмоционального контекста фрагментов текста или текстовых документов на естественном языке. Рассмотрены факторы, влияющие на анализ эмоционального контекста, делается обзор алгоритмов измерения тональности текста. Предлагается новый комплексный подход для выявления эмоционального контекста, основанный на объединении различных методов. Обсуждаются потенциальные перспективы применения предложенного метода и его влияние на развитие анализа текстов с учетом эмоциональной окраски.

Ключевые слова. Анализ текста, тональность текста, эмотивность, контекст.

Введение. Исследования, направленные на выявление эмоций человека через анализ устных и письменных высказываний, являются сегодня очень актуальными, хотя и сложными, многогранными. Лингвисты стараются систематизировать языковые единицы, которые выражают или обозначают эмоции, однако различные авторы придают этому понятию разные значения. Это связано с «расплывчатости самого концепта, принадлежащего ряду смежных с лингвистикой наук» [1].

Сложности в исследовании и описании текстового выражения эмоциональных состояний возникают из-за неоднозначности понятия эмоции. Многие учёные считают, что эмоциональное состояние является отклонением от нейтрального, условного нуля, хотя некоторые из них не признают этот нуль абсолютным, предпочитая говорить о относительно нейтральном эмоциональном состоянии. В настоящее время в психологии эмоций нет единой теории эмоциональных явлений, что приводит к различным интерпретациям термина «эмоция».

Сегодня исследователи, которые занимаются анализом текста и, в частности, выявлением эмоционального контекста, относятся к одному из двух направлений – лингвисты или математики. При этом у обеих сторон прослеживается ряд проблем. Математические методы показывают стабильный и быстрый результат на больших выборках данных, но, в силу своей неповоротливости, зная алгоритм его легко можно сломать. Рассмотрим пример, попробовав обработать фразу: «да нет, мне ужасно понравилось». Алгоритм обратится к заранее подготовленному тональному словарю и увидит два слова с флагом «положительно» и два с «отрицательно», после чего можно сделать логичный (по алгоритму), что результат – эмоция нейтральная. Но это далеко не всегда так.

При лингвистическом подходе к оценке эмоциональной тональности материалов необходимо хорошо разбираться в языковых нормах, уметь выражать эмоции, чувства и действия, а также определять их. Во-вторых, для выявления эмоциональной окраски текста требуется быть внимательным и наблюдательным, отмечать детали описываемых событий и их характеристики. В-третьих, эмоциональный анализ включает как проект в целом, так и его отдельные части: конкретные объекты, действия, результаты, общую картину или позицию автора. Важно объективно воспринимать описываемые явления, правильно их интерпретировать и объяснять. Проще говоря, трудность заключается в том, что лингвистический анализ текста сможет провести только специально обученный человек. На работу он потратит большое количество времени. А если еще и учитывать человеческий фактор, самочувствие человека, какие-то личные взгляды на вещи и так далее – о верном и конструктивном результате можно забыть.

Основная часть. Если брать текущее состояние решения вопроса, то математики в основном располагают методами автоматического определения эмоциональной составляющей (тональности) в тексте и описывают опыт осуществляемой в данный момент практической реализации системы для текстов на естественном языке, в основе которой лежат словари лексической тональности и набор комбинаторных правил объединения отдельных слов и словосочетаний [2].

Известны три основных метода определения тональности текста:

– анализ текста методами векторного анализа (часто с применением n-граммных моделей), сравнение с ранее размеченным эталонным корпусом по выбранной мере близости и отнесение (классификация) текста к негативу или позитиву на основании полученного результата сравнения;

– поиск эмотивной лексики (лексической тональности) в тексте по заранее составленным тональным словарям (спискам паттернов) с применением лингвистического анализа. По совокупности найденной эмотивной лексики текст может быть оценен по шкале, отражающей количество негативной и позитивной лексики. Этот метод может использовать как списки паттернов, подставляемые в регулярные выражения, так и правила соединения тональной лексики внутри предложения.

– смешанный метод (комбинация первого и второго подходов).

Наиболее эффективным и распространённым методом анализа текста является векторный анализ, который включает такие подходы, как «наивный байесовский классификатор», «случайный лес», «рекуррентная нейронная сеть» и другие. В векторном анализе основным показателем успешности считается оценка тональности с точностью около 70-80%, что обычно достигается при анализе естественного языка. Среди перечисленных методов наиболее результативным является «наивный байесовский классификатор». Этот классификатор базируется на алгоритме байесовского классификатора, где для каждого классифицируемого объекта вычисляются функции правдоподобия для всех классов. На основе этих значений определяются апостериорные вероятности классов, и объект относится к тому классу, для которого апостериорная вероятность максимальна. Для работы модели вводится допущение о независимых признаках [3].

По результатам анализа можно определить эмоциональную окраску текста: положительную, нейтральную или отрицательную. Положительная эмоциональная окраска указывает на то, что текст вызывает положительные эмоции, такие как радость, восторг или удовлетворение. Например, описание приятного события или хорошей новости будет иметь положительную эмоциональную окраску. Отрицательная эмоциональная окраска означает, что текст вызывает отрицательные эмоции, такие как грусть, разочарование или злость. Тексты, рассказывающие о трагических событиях или плохих новостях, обычно обладают отрицательной эмоциональной окраской. Нейтральная эмоциональная окраска означает, что текст не вызывает ярко выраженных эмоций и является информативным или объективным, не влияя на эмоциональное состояние читателя. Примерами нейтральной окраски могут быть тексты, содержащие описание фактов или научных данных. [4].

Лингвисты же, в свою очередь, не останавливаются на одном определении «положительный» либо «отрицательный» окрас. С лингвистической точки зрения эмоциональный анализ – это последовательный и насыщенный алгоритм, позволяющий уловить не только общие «краски», логику изложения мыслей, но и ракурс исследования, принадлежность к определенному жанру, выделить специфические черты и манеры изречения и прочее. В данном случае объектом внимания эксперта становятся именно слова и выражения, подчеркивающие состояние героев, авторскую позицию, общий сюжет и характер произведения в целом [6]. Ключевые моменты эмоционального анализа текста представлены в таблице 1.

Таблица 1 – Ключевые моменты эмоционального анализа текста

<i>Тональность текста</i>	<i>Строение текста</i>	<i>Авторская позиция</i>
Общий характер проекта	Специфические конструкции	Обоснованность и убедительность утверждений, выводов
Впечатления от прочтения	Использование определенных или неординарных слов, фраз	Отношение автора к теме
Эмоции и средства выражения		

Реализация оценки эмоциональной тональности материалов предполагает следование определенной схеме. Диагностика и определение ключевых моментов начинаются с простой вычитки проекта или произведения. Схема проведения эмоционального анализа текста представлена на рисунке 1.

На первом этапе происходит изучение предоставленных материалов для получения общего представления о содержании текста. Эксперт должен понять тему текста, определить его общий характер, отметить стиль речи, границы исследования и другие важные аспекты.

Второй этап эмоционального анализа включает проведение общей стилистической экспертизы. Здесь важно выделить и обосновать стиль изложения, а также отметить наличие или отсутствие средств выражения, подчеркивающих эмоциональную окраску текста. В каждом жанре существуют свои эмоциональные приёмы и средства выражения. Необходимо выделить их в анализируемом материале и показать их принадлежность к конкретной стилистической группе, соответствие языка повествования стилистическим нормам и требованиям или их противоречие.

Третий шаг – лексический анализ. Этот этап крайне важен для оценки тональности текста. Он предполагает детальное рассмотрение каждого слова и фразы, их предназначения, роли в тексте, уместности в логическом и стилистическом контексте. Лексический анализ позволяет выделить эмоционально окрашенные категории и оценить корректность, необходимость и возможности их использования в общем контексте материала.

Четвёртый шаг – синтаксический анализ текста. Эмоции в текстах выражаются не только через отдельные слова и конструкции, но и через особенности построения речи. Синтаксический анализ позволяет определить точное значение и характер выражений, эмоций, а также их место в общем контексте, их предназначение и суть.

Пятый шаг – определение морфологических особенностей интонаций и общей тональности текста. Выделяя части речи и их сочетания, можно определить характер выражений (удивление, восклицание, вопрос, повествование, констатация, убеждение и т.д.). Морфологический анализ также помогает определить читабельность и восприимчивость материала.

Шестой шаг – обобщение результатов анализа и подведение итогов относительно эмоциональной составляющей произведения, а также соответствия тональности стилистическим и лингвистическим параметрам.

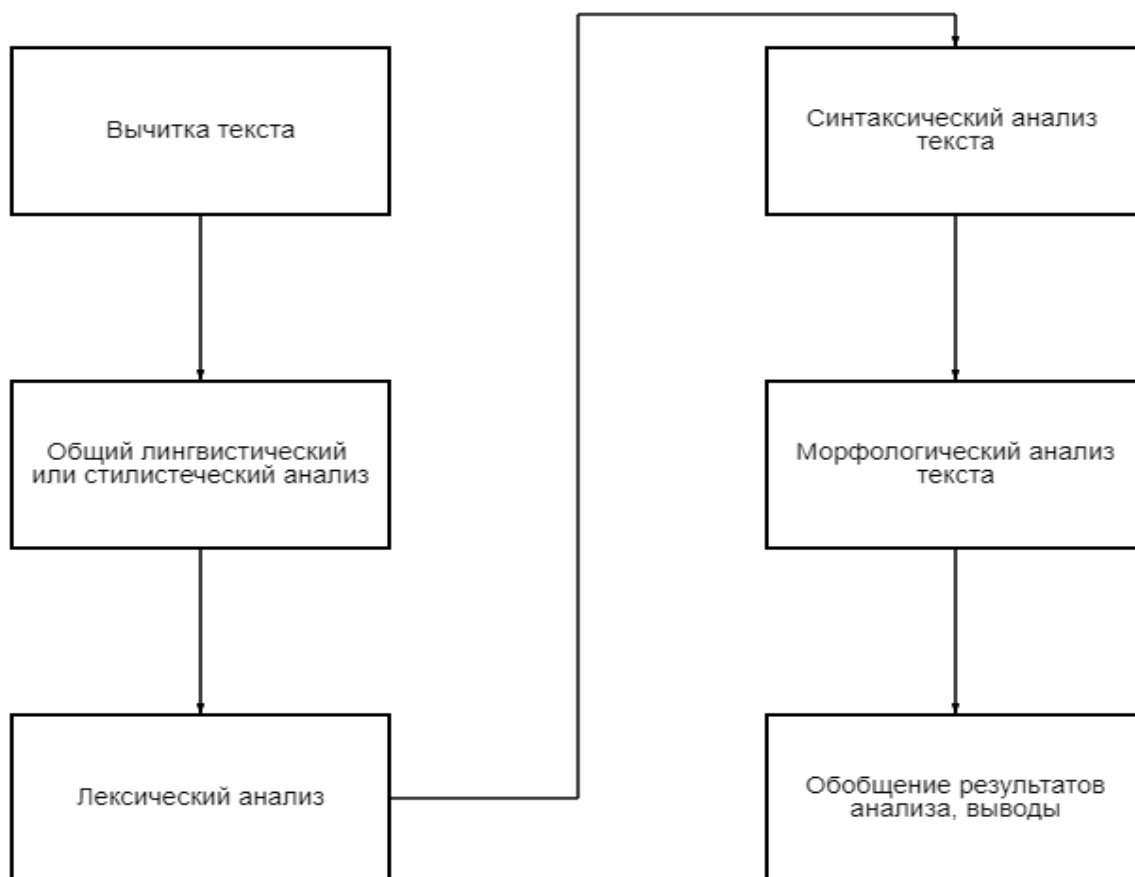


Рисунок 1 – Схема проведения эмоционального анализа текста

Посредством комплексного анализа письменного художественного текста, содержащего описание эмоционального состояния субъекта, можно показать фрагмент языковой картины мира, воплощенный в художественном тексте.

Для этого используются методы лингвистического описания и комплексного анализа текста, сравнительный метод, метод моделирования, анализ и систематизация фактического материала.

Так же лингвисты рассматривают анализ текста со стороны социальной коммуникации. Социальная коммуникация есть движение знаний, эмоциональных переживаний, волевых воздействий в социальном времени и пространстве.

Чтобы получить вещественный результат, необходимо обработать и транслировать лингвистический подход на математическую область. Концепт переноса лингвистических методов может включать в себя автоматизацию эмоционального анализа, что значительно уменьшит время обработки. Благодаря автоматизации и цифровизации эмоционального анализа есть возможность интерпретировать значения результатов в численный вид.

Получив численные данные, перед нами открывается обширный пласт возможностей по использованию математических методов анализа. К примеру, результат вычислений можно представить в виде координатной оси, где отклонение от логического нуля (нейтрального состояния) будет означать отклонение в сторону радости/грусти, злости/сочувствия и так далее.

По числовым данным открывается возможность рассчитать дисперсию, корреляцию и так далее. Высокая дисперсия может указывать на разнообразие эмоциональных реакций в текстах или

коммуникациях. Например, если дисперсия оценок эмоциональной окраски высока, это может означать, что в данных присутствует широкий спектр эмоций, от радости до грусти.

Анализ корреляции между различными эмоциональными состояниями или между эмоциональным состоянием и другими переменными может помочь выявить связи между ними. Например, можно анализировать корреляцию между эмоциональным состоянием и содержанием текста, чтобы выяснить, какие эмоции чаще возникают при обсуждении определенных тем или событий. Например, можно исследовать, какие эмоции наиболее часто проявляются в отзывах о фильмах или политических дебатах, и выявить особенности восприятия различных тем обсуждений.

Заключение. Исследование, рассматривающее факторы, влияющие на анализ эмоционального контекста текста, представляет значительный вклад в развитие области анализа текстов с учетом их эмоциональной окраски. Новый подход, основанный на объединении различных методов, предоставляет перспективу для улучшения точности и эффективности анализа.

Важно отметить, что область применения данного исследования широка и разнообразна. Оно может быть полезным для различных сфер, таких как мониторинг общественного мнения в социальных сетях, анализ тональности отзывов о товарах и услугах, оценка эмоциональной составляющей текстов в новостных изданиях и медиа, а также для поддержки принятия решений в маркетинге и рекламе. Исследование в этой области может способствовать созданию более точных и автоматизированных инструментов для анализа текстов, что позволит компаниям и организациям получать более надежные и информативные данные о восприятии их продуктов или услуг.

Кроме того, развитие методов анализа эмоционального контекста может способствовать улучшению понимания социальных и культурных тенденций, а также помочь в выявлении настроений и трендов в обществе. Таким образом, исследование в этой области имеет потенциал для создания новых инструментов и подходов, которые могут быть применены в различных сферах деятельности, что в свою очередь способствует развитию и улучшению анализа текстов с учетом их эмоциональной окраски.

Список использованных источников:

1. Моделирование как метод комплексного анализа эмотивного текста [Электронный ресурс] – Режим доступа: <https://science-education.ru/ru/article/view?id=15872>. – Дата доступа : 22.03.20024
2. Метод определения эмоций в текстах на русском языке [Электронный ресурс] – Режим доступа: https://masters.donntu.ru/2020/fknt/pilipenko/library/t_a_ricle3.htm. – Дата доступа : 21.03.20024
3. Сентимент-анализ [Электронный ресурс] – Режим доступа: <https://habr.com/ru/companies/simbirsoft/articles/539508/>. – Дата доступа : 12.03.20024
4. Как нейронные сети помогают анализировать эмоциональную окраску текста [Электронный ресурс] – Режим доступа: <https://nauchniestati.ru/spravka/analiz-emocionalnoj-okraski-generiruemogo-teksta-na-osnove-nejronnyh-setej/>. – Дата доступа : 22.03.20024
5. Порядок и правила проведения эмоционального анализа текста [Электронный ресурс] – Режим доступа: <https://disshelp.ru/blog/poryadok-i-pravila-provedeniya-emotsionalnogo-analiza-teksta/>. – Дата доступа : 24.03.20024
6. Введение в теорию социальной коммуникации [Электронный документ] – Режим доступа: http://library.lgaki.info:404/2017/Соколов%20А_Введение%20в%20теорию.pdf. – Дата доступа : 22.03.20024

UDC 004.9+81

METHODS AND MODELS OF TEXT ANALYSIS TO IDENTIFY EMOTIONAL CONTEXT

Paulichenka K.A.

*Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Paramonov A.I. – Candidate of Engineering Sciences, Associate Professor

Annotation. In this scientific work, the problem of identifying the emotional context of fragments of text or text documents in natural language is investigated. The factors influencing the analysis of the emotional context are considered, and the algorithms for measuring the tonality of the text are reviewed. A new comprehensive approach is proposed to identify the emotional context based on combining various methods. The potential prospects for the application of the proposed method and its impact on the development of text analysis, taking into account emotional coloring, are discussed.

Keywords. Text analysis, text tonality, emotivity, context.