

УДК 378+004.93

## АНАЛИЗ ЭМОТИВНОСТИ СООБЩЕНИЙ КАК ИНСТРУМЕНТ АДАПТИВНОГО ОБРАЗОВАНИЯ

Куличок Н.С., Парамонов А.И.

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

[nikita.kulicok@gmail.com](mailto:nikita.kulicok@gmail.com), [a.paramonov@bsuir.by](mailto:a.paramonov@bsuir.by)

Рассматривается подход к применению анализа эмоциональных характеристик текста для реализации программы адаптивного обучения. Задача анализа эмоциональной составляющей текста сводится к общей проблеме анализа цифровых текстов на естественном языке и дальнейшему выявлению их эмотивных характеристик. Дается понятие эмотивности и ее составляющих. Рассмотрены существующие методы анализа тональности текстов как основы эмотивного фона текста. Проведено сравнение точности этих методов.

Ключевые слова: адаптивное образование; эмотивность текста; анализ тональности; машинное обучение.

Адаптивное образование – подход к образованию, позволяющий максимально учитывать индивидуальные особенности учащихся. Адаптивное образование может быть представлено различными способами. Например, как предоставление учащимся ссылок на дополнительные материалы по наиболее интересным им темам. Или рекомендация учащемуся обучающих курсов на основе данных об уже изученных им курсах. Так или иначе, адаптивное обучение предполагает некоторый анализ данных об учащемся, таких как пройденные им курсы, его отношение к ним, его интересы для определения пути дальнейшего обучения. Как инструмент для выполнения данной задачи может быть использован анализ эмотивности данных об интересах учащихся, их ответах и отзывах, оценках.

Свойство языка выражать эмоциональные состояния и переживания человека принято называть эмотивностью. В речи функция эмотивности реализуется с помощью специальных средств языка – эмотивов. Эмотивы представлены в любом естественном языке и образуют его так называемый эмотивный код. Таким образом, эмотивный код языка – это система корреляций между эмоциями тех, кто взаимодействует, и их концептуализацией. В свою очередь эмотивный код текста можно представить как систему некоторых характеристик, которые отражают общую эмоциональную окраску текста, а также отношение автора к описываемому в нём.

Существует множество характеристик эмотивности, которые можно анализировать, в числе которых: эмотивный фон, эмотивная тональность, эмотивная окраска, эмотивная направленность, эмотивная модальность, эмотивные интенции текста [2]. Наибольшего успеха исследования эмотивности текста достигли при анализе его тональности. Анализ тональности – класс методов контент-анализа в компьютерной лингвистике, предназначенный для автоматизированного выявления в текстах эмоционально окрашенной лексики и эмоциональной оценки авторов по отношению к объектам, речь о которых идёт в тексте [3].

Для определения настроения в анализе тональности используются три термина: объект, о котором дается мнение, особенности этого объекта, а также владелец мнения об объекте [4]. Анализ тональности может выполнять задачу классификации на трех уровнях: на уровне документа, уровне предложения и уровне характеристик. Уровень классификации документа используется там, где задача состоит в нахождении общей тональности темы независимо от того, кто именно придерживается мнения. Классификация по уровням предложений предполагает, что каждое предложение придерживается единого мнения. На уровне характеристик (или аспектов) выполняется анализ различных характеристик объекта. Анализ тональности включает в себя предварительную обработку данных, выбор характеристик и классификацию, а затем определение тональности данных. Предварительная обработка данных включает в себя токенизацию, удаление стоп-слов, выделение корней, лемматизацию и другие преобразования.

Методы анализа тональности можно поделить на две большие группы: основанные на правилах и словарях и основанные на машинном обучении.

Методы на основе правил и словарей включают две соответствующие подгруппы: методы на основе правил и методы на основе словарей. В основе методов на основе правил лежит идея, что система текста состоит из некоторого набора правил, применяя которые и делается заключение о тональности текста. В данном случае для достижения хорошей точности при анализе необходимо составить большое количество правил. Зачастую правила привязаны к определенной предметной области и при её изменении потребуется заново составлять их. Тем не менее, при наличии достаточного количества правил данный подход является одним из наиболее точных. Методы на основе словарей используют так называемые тональные словари, которые представляют собой списки слов с указанным для каждого некоторого значения его тональности. Как правило, это значение представляет собой некоторый коэффициент. Для достижения хорошей точности при анализе необходимо вычислить значения двух оценок: положительной составляющей текста и отрицательной. Положительная составляющая текста вычисляется как сумма тональностей всех положительных терминов, которые присутствуют в тексте. Аналогичным образом рассчитывается значение отрицательной составляющей. Итоговая оценка тональности всего исследуемого текста рассчитывается как отношение этих двух составляющих. Естественно, точность результатов напрямую зависит от размера и качества словаря, то есть корректности указанных коэффициентов тональности слов. Построение словарей является отдельной научной задачей.

В методах машинного обучения для анализа используется помеченный набор данных, где тональность предложений уже определена. На основе этого набора данных извлекается признак, который затем помогает классифицировать точность неизвестного входного предложения. В свою очередь методы машинного обучения можно разделить на 2 группы: на обучение с учителем и обучение без учителя. Машинное обучение с учителем используется в случае, когда для обучения модели доступны помеченные данные [5]. Реализация этого метода предполагает выполнение двух этапов: первый шаг – обучение модели, второй – решение задачи прогнозирования. Во время обучения набор данных с его метками подается в алгоритм классификации, который на их основе формирует модель в качестве выходных данных. После этого данные, тональность которых необходимо определить, вводятся в модель для прогнозирования их категории. Сегодня известны различные алгоритмы классификации для машинного обучения с учителем, такие как наивный байесовский классификатор, метод опорных векторов, нейронные сети и другие. Метод машинного обучения без учителя используется, когда сбор помеченных данных затруднен. Зачастую собрать непомеченные данные легче, чем помеченные. При таком подходе предложения и документы классифицируются на основе списков ключевых слов каждой категории, которые также должны быть заранее подготовлены.

Для задачи анализа тональности текстов уже существует множество наборов помеченных данных. Поэтому методы машинного обучения с учителем являются более предпочтительными.

Наивный байесовский классификатор [6] – это семейство простых вероятностных классификаторов, основанных на использовании теоремы Байеса и предположении о независимости признаков классифицируемых объектов. Алгоритм является масштабируемым по числу признаков, а по точности сопоставим с другими популярными методами, такими как машины опорных векторов. Как и любой классификатор, байесовский присваивает метки классов объектам, представленным векторами некоторых признаков. При этом предполагается, что каждый признак независимо влияет на вероятность принадлежности объекта к классу. Простой байесовский классификатор строится на основе обучения с учителем. Дополнительным преимуществом метода является небольшое число данных, которые необходимы для обучения.

Метод опорных векторов [7] – семейство алгоритмов бинарной классификации, основанных на обучении с учителем, использующих линейное разделение пространства

признаков с помощью гиперплоскости. Основная идея метода заключается в отображении векторов пространства признаков, представляющих классифицируемые объекты, в пространство более высокой размерности. После перевода в пространство большей размерности, в нём строится разделяющая гиперплоскость. При этом все векторы, расположенные с одной стороны гиперплоскости, будут относиться к одному классу, а расположенные с другой – к другому. Также по обе стороны основной разделяющей гиперплоскости, параллельно ей и на равном расстоянии от неё строятся две вспомогательные гиперплоскости, расстояние между которыми называют зазором. Задача заключается в построении разделяющей гиперплоскости таким образом, чтобы максимизировать зазор – область пространства признаков между вспомогательными гиперплоскостями, в которой не должно быть векторов. Векторы, которые попадут на границы зазора, то есть будут находиться на вспомогательных гиперплоскостях, и будут называться опорными векторами.

Нейронная сеть [8] – структура, которая состоит из искусственных нейронов, которые определенным образом связаны друг с другом и внешней средой с помощью связей, каждая из которых имеет определённый коэффициент, на который умножается поступающее через него значение. Данный коэффициент называется весом. Нейронные сети могут быть реализованы как программно, так и аппаратно. В процессе функционирования нейронная сеть осуществляет преобразование данных, конкретный вид которого определяется весами связей, видом активационной функции нейронов, а также архитектурой и конфигурацией сети. Нейронные сети представляют собой модели, основанные на машинном обучении. Модели формируются в процессе обучения. Обучение заключается в итеративной подстройке весов нейронной сети под нужный результат. Подстройка основана на некотором правиле, которое называется алгоритмом обучения.

В результате исследования был произведён анализ точности данных методов. Результат представлен в таблице 1.

Таблица 1 – Точность методов анализа тональности

Метод	Точность
Наивный байесовский классификатор	64%
Метод опорных векторов	78%
Нейронные сети	80%

Таким образом, использование методов анализа тональности позволит эффективнее обрабатывать данные об учащихся для определения дальнейшего пути их обучения при адаптивном образовании. В настоящий момент данная задача является актуальной в связи с развитием новых подходов к образованию, таких как различные онлайн-курсы.

### Литература

1. Пазельская А., Соловьев А. Метод определения эмоций в текстах на русском языке // The international conference on computational linguistics and intellectual technologies “Dialogue 2011” : конференция. – Москва, 2011. – 12 с.
2. Ленько Г.Н. Уровни анализа текстовой эмотивности (на примере текстов художественного стиля) // Вестник ЛГУ им. А.С. Пушкина. – 2014. – №2. – 10 с.
3. Семина Т.А. Анализ тональности текста: современные подходы и существующие проблемы // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6, Языкознание: Реферативный журнал. – 2020. – №4. – 18 с.
4. Алгоритмы анализа тональности текста / Н. С. Иванин и другие // BIG DATA and Advanced Analytics: collection of materials of the third international scientific and practical conference. (Minsk, Belarus, May 3–4, 2017) / editorial board : М. Batura [etc.]. – Minsk : BSUIR, 2017. – 5 с.
5. Самигулин Т.Р., Джурабаев А.Э.У. Анализ тональности текста методами машинного обучения // Научный результат. Информационные технологии. – Т.6. – №1. – 2021. – 7 с.
6. Турканов Г.И., Щепин Е.В. Классификатор Байеса для переменного количества признаков // Труды МФТИ. – 2016. – №4 (32). – 4 с.
7. Федотов Д.В. О решении задачи классификации методом опорных векторов // Решетневские чтения. 2013. №17. – 3 с.
8. Ксенофонтов В.В. Нейронные сети // Проблемы науки. – 2020. – №11 (59). – 2 с.

IV МНПК «Непрерывное профессиональное образование лиц с особыми потребностями»

## **EMOTIONALITY ANALYSIS OF MESSAGES AS AN ADAPTIVE LEARNING TOOL**

Kulichok N.S., Paramonov A.I.

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

The approach to the application of the text emotional characteristics analysis for the adaptive learning program implementation is considered. The task of analyzing the text emotional component is reduced to the general problem of analyzing digital texts in natural language, and further identifying their emotive characteristics. The concept of emotivity and its components is given. The existing methods of the text's tonality analysis as the basis of the text emotive background are considered. The accuracy of the considered methods is compared.

Keywords: adaptive learning; emotive text; sentiment analysis; machine learning.