

МЕТОДЫ ОЦЕНКИ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ ЛЕКСИКИ

Юань Юе

Аспирант, Московский государственный университет
имени М.В. Ломоносова
853342287@qq.com

FACTORS INFLUENCING THE EMOTIONAL COLORING OF VOCERATORS

Yuan Yue

Summary: Within the framework of this article, an analysis of the methods that are used to create the emotional coloring of the text is carried out, the work also examines the internal and external factors that affect the emotional color of the vocabulary. The relevance of this study is due to the fact that at the moment there is a constantly growing interest in the issues of the emotional coloring of the text, in the emotional component of the text in general, as well as the increasing interest of researchers in the analysis of the emotional component of the text. The article discusses current problems and suggests ways to solve them.

Keywords: analysis, methods, emotional coloring, text, study.

Аннотация: В рамках данной статьи проведен анализ методов, которые применяются для создания эмоциональной окраски текста, также в работе рассматриваются внутренние и внешние факторы, которые влияют на эмоциональную окраску лексики. Актуальность данного исследования обусловлена тем, что на данный момент постоянно повышается интерес к вопросам эмоциональной окраски текста, к эмоциональной составляющей текста вообще, а также увеличением заинтересованности исследователей в анализе эмоциональной составляющей текста. В статье рассматриваются актуальные проблемы и предлагаются способы их решения.

Ключевые слова: анализ, методы, эмоциональная окраска, текст, изучение.

Анализ эмоциональной окраски текста позволяет определить мнения и аффективные намерения авторов данных текстов, а также их отношения, оценки и склонности по отношению к разным темам. На сегодняшний день известны исследования, посвященные анализу эмоциональной составляющей в различных текстовых жанрах, включая обзоры продуктов и фильмов, новости, редакционные статьи и статьи-мнения, а также текстов из блогов [1, 2, 3]. Работа над анализом эмоциональной окраски текста обычно сосредоточена на распознавании эмоциональной валентности – положительна она или отрицательна. Среди менее изученных областей эмоциональной составляющей – распознавание типов эмоций, их силы и интенсивности. Исследование эмоций в целом в последнее время привлекает повышенное внимание научного сообщества.

Целью данного исследования является рассмотрение методов определения эмоциональной окраски лексики текста.

Для достижения поставленной цели использовались такие методы сравнительный анализ, систематизированный литературный анализ, контент-анализ, формализация. В качестве материалов использовались публикации зарубежных и отечественных авторов, посвященные эмоциональной тональности текстов.

На данный момент активными темпами развивается такое перспективное направление в лингвистике как сентимент-анализ, он же анализ эмоциональной тональности. Ключевой задачей данного направления является процесс нахождения в тексте лексики, которая имеет эмоциональную окраску, и оценку данных лексических

значений для формирования представления об эмоциональном тоне всего анализируемого текста.

Эмоциональную окраску, которая присутствует в тексте, принято называть его тональностью [4]. Она представляет собой эмоциональную составляющую, которая в тексте выражена набором специализированных лексических средства [5]. То, какая у текста эмоциональная окраска, можно определить при помощи тональности его составляющих и взаимосвязей между ними.

На основании выполненного систематического литературного анализа было выявлено, что эмоциональная окраска текста может быть 3 основных типов:

- бинарной (положительной и отрицательной),
- ринарной (положительной, отрицательной и нейтральной),
- ранжированной.

На основании приведенной классификации можно сделать вывод, что в оценку может входить помимо валентности еще и интенсивность высказывания. Выведение интегральной оценки может быть основано на ряде характеристик [6] или же на общей тональности. Сложно осуществить определение эмоциональной окраски целого текста ввиду часто различающихся оценок, по этой причине гораздо эффективнее оценивать каждое предложение по отдельности, в ряде случаев требуется разбивать предложение на отдельные словосочетания, фразы и другие единицы [7].

Рассматривая методы для оценки эмоциональной окраски лексики текста, необходимо определить, что такое авторские оценочные суждения и как реализуются

ся их оценка в тексте. Авторские оценочные суждения представляют собой совокупность, включающую в себя валентность, объект-субъект. Объект в данном случае – это то, в отношении чего непосредственно реализовано эмоциональное высказывание. Субъект – это тот, от чьего лица исходит высказывание, а валентность – это то, как автор высказывания относится к объекту с эмоциональной точки зрения [8]. Перед тем как определять каждый из рассмотренных выше компонентов, требуется выделить оценочные мнения.

Для того чтобы разграничить тексты, которые содержат в себе оценочные суждения, и те, в которых реализуется просто описание фактов, принято использовать стандартные методы классификации, которые перечислены ниже.

Одним из часто применяемых методов классификации текстов на основании определения эмоциональной окраски является проведение анализа словосочетаний. Для использования данного метода производится синтаксическая разметка текста. Далее, базирясь на существующих шаблонах, производят выделение словосочетаний, которые, скорее всего, имеют оценочный характер. Далее осуществляется определение эмоциональной валентности, а итоговой оценкой будет являться соотношение полученных оценок друг с другом. Словосочетания позволяют сделать вывод о предложениях.

Для формирования вывода о словосочетаниях как единицах эмоционально окрашенной лексики, выполняется анализ отдельно взятых слов, биграмм и тиграмм, из которых состоят анализируемые словосочетания. В ходе анализа главной задачей является оценить, насколько точны все перечисленные элементы.

Простейшая классификация не дает возможности произвести учет взаимосвязей, по этой причине существуют точки зрения [9], согласно которым осуществляется применение графового подхода, согласно которому каждая вершина является предложением.

Анализ эмоциональной окраски текста принято проводить с использованием средств машинного обучения, что существенно повышает точность и достоверность полученных результатов. Признаками могут выступать методы оценки текстовой эмоциональной окраски, слова или биграммы.

Как правило, каждый текст включают в себя много информации и лексики, которые лишены эмоциональной окраски, то есть не могут стать базой или материалом для выполнения оценки эмоциональной тональности. Чаще всего системы реферирования базируются на познаниях в части предметной области или же на глубоком анализе с точки зрения лингвистики. Особого

внимания заслуживает система под названием Opinosis. В ней текст представлен в формате ориентированного графа. Его вершиной выступает словосочетание, а ребра демонстрируют структуру предложений. Такая реализация дает возможность распознать информацию общего и различного характера.

Крайне важно адаптировать классификаторы мнений к имеющимся предметным областям. Такая проблема может быть решена несколькими способами:

- реализация обучения классификаторов параллельно на всех существующих информационных наборах;
- разграничение признаков, которые используются на разных доменах, то есть для каждого из них должен быть создан специальный словарь.

Это даст возможность исключить специфические высказывания, но сохранить эмоционально окрашенные тексты независимо от области [10].

Сложность в проведении исследований эмоциональной окраски текста обусловлена проблематикой выделения единиц оценки тональности, а также тем, что эмоциональная составляющая лексических компонентов крайне неоднозначна. Так, например, несмотря на одну предметную область, «высокая стоимость» является отрицательным показателем товарной единицы, а вот «высокое качество» – наоборот является положительной характеристикой товарной единицы.

Для анализа тональности текста активно применяются статистические методы, потому что они отличаются простотой, однако сложность сопряжена с тем, что нужен размеченный корпус, а это трудоемко и не дешево [11].

Таким образом, методы, которые базируются на поиске эмотивной лексики в тональных словарях, сопряжены с зависимостью от полноты представленной лексики. Необходимо учитывать, что рассматриваемые подходы имеют предметную зависимость, то есть для разных предметных областей необходимо подготавливать отдельные словари. Анализ эмоциональной окраски лексики концентрируется на символическом уровне текстов в рамках прикладных исследований. Метод основан на изучении ассоциации слов, словосочетаний, передающих большую часть эмоциональных компонентов текстов. В рамках данного подхода язык рассматривается как организатор отношений между автором текста и контекстом, а не как детектор эмоций человека. Данный подход отличается от аналитической работы, изучающей грамматические, синтаксические и семантические характеристики текста, тем, что оценивает эмоциональные категории: валентность, интенсивность, силу.

ЛИТЕРАТУРА

1. Pang B., Lee L., Vaithyanathan S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proc. Conf. on EMNLP – 2002 – p. 115-126.
2. Turney P.D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proc. the 40th Annual Meeting of the ACL, Philadelphia – 2002 – P. 245.
3. Wiebe J., Wilson T., Cardie C.: Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 39(2-3) – 2005—p. 165–210.
4. Gamon M., et al. Pulse: Mining Customer Opinions from Free Text // Proceedings of the 6th International Symposium on Intelligent Data Analysis (IDA). 2015. P.121–132.
5. Wiebe J., Wilson T., Cardie C. Annotating Expressions of Opinions and Emotions in Language // Proceedings of Language
6. Пазельская А., Соловьев А. Метод определения эмоций в текстах на русском языке: труды международной конференции «Диалог, 2021». P.510– 522.
7. «Дорожки РОМИП'2021» Available: <http://romip.ru/ru/2021/tracks.html>. [Дата обращения: 15.07.2021].
8. Ng R., Pauls A. Multi-document summarization of evaluative text // Proceedings of the 11st Conference of the European Chapter of the Association for Computational Linguistics. 2016. P.305– 312.
9. Carenini G., Ng R. Zwart E. Extracting knowledge from evaluative text // Proceedings of the 3rd international conference on Knowledge capture. 2015. P.11–18.
10. Turney P. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews // Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. 2017. P. 417–424.
11. Wiebe J., Bruce R., O'Hara T. Development and use of a gold-standard data set for subjectivity classifications // Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. 2019. P. 246–253.

© Юань Юе (853342287@qq.com).

Журнал «Современная наука: актуальные проблемы теории и практики»



Московский государственный университет имени М.В. Ломоносова