

Классификация тональности мнений для задачи автоматического сентимент-анализа текста

Чернышевич М.В.

Учреждение образования «Минский государственный лингвистический университет», Минск

В работе исследуется задача классификации тональности мнений, решение которой может быть положено в основу построения актуальной системы автоматического сентимент-анализа текста (АСАТ).

Цель статьи – проанализировать основные существующие типы классификаций тональности мнений, их достоинства и недостатки, разработать новую, ориентированную на массового пользователя, шкалу тональности мнений, оперирующую как абсолютными, так и сравнительными оценками.

***Материал и методы.** Материал исследования – корпус сообщений пользователей, отобранных методом случайной выборки из трех источников: социальной сети twitter.com, социальной сети facebook.com, а также интернет-платформ fixya.com, amazon.com и tripadvisor.com. Метод – контекстуальный анализ.*

***Результаты и их обсуждение.** Большинство существующих решений задачи АСАТ ориентировано на бинарную шкалу тональности (положительная и отрицательная категории тональности), иногда к ней добавляется категория «нейтральная тональность», а также количественная шкала для любой из используемых категорий. Но массовым пользователем сейчас очень востребованы более гибкие шкалы, ориентированные, в частности, на решение задачи на уровне объектов и аспектов, а не предложений и тем более документов в целом, на применение наряду с общими и частотных оценок, содержащих в том числе и дескриптивные значения, указывающие на основание оценки, а также сравнительных оценок. Опираясь на анализ исследуемого текстового материала, актуальность получаемых по ходу исследования практических результатов для массового пользователя, разработана шкала тональности мнений, оперирующая как абсолютными (общими и частотными), так и сравнительными оценками. Дано формальное определение мнения. Для общих оценок предложена тональность типа Positive, Negative и Desire, для частотных – Appreciation, Deficiency и Wish, для сравнительных – ComparisonPositive, ComparisonNegative и ComparisonNeutral.*

***Заключение.** Следовательно, разработанная и представленная в статье классификация тональности мнений является одной из наиболее приемлемых для современного и качественного решения задачи автоматического сентимент-анализа текста.*

***Ключевые слова:** сентимент-анализ текста, классификация тональности мнений, тональность мнения.*

Opinion Classification for Automatic Sentiment Analysis of the Text

Chernyshevich M.

Educational Establishment «Minsk State Linguistic University», Minsk

The article deals with opinion sentiment classification that is the basis for building a modern automatic sentiment analysis system.

The purpose of the work is to analyze the main existing types of opinion sentiment classifications, their advantages and disadvantages, to develop a new, user-focused, sentiment scale, that deals with both absolute and comparative opinions.

***Material and methods.** The research material is a corpus of random user messages from the following sources: the social networks twitter.com and facebook.com, the Internet platforms fixya.com, amazon.com and tripadvisor.com. The research method is contextual analysis.*

***Findings and their discussion.** Most of the existing solutions of the sentiment-analysis problem are focused on the binary sentiment classification (positive and negative categories), sometimes the category «neutral sentiment» is added, as well as the quantitative scale for any of the used categories. But the user today greatly demands for more flexible scales, oriented to solve the problem at the level of objects and*

aspects, rather than sentences and documents, to recognize not only general opinions, but also descriptive and comparative opinions. A novel sentiment scale of opinions was developed based on the analysis of the research material and the relevance of the practical results for the mass user. The proposed scale deals with both absolute (general and descriptive) and comparative opinions. A formal definition of opinion is given. Sentiment categories Positive, Negative and Desire are proposed for general opinions, Appreciation, Deficiency and Wish – for descriptive opinions, and ComparisonPositive, ComparisonNegative and ComparisonNeutral – for comparative opinions.

Conclusion. *The classification of sentiments developed and presented in the article is one of the most suitable scales for a modern and high-quality automatic sentiment analysis system.*

Key words: *sentiment analysis, opinion classification, sentiment classification, opinion polarity.*