

ИЗВЛЕЧЕНИЕ ЭМОЦИЙ ИЗ РУССКОЯЗЫЧНЫХ ТЕКСТОВ

Автор: Слапогузов А.П., Университет ИТМО, Санкт-Петербург

Научный руководитель: Цопа Е. А., Университет ИТМО, Санкт-Петербург

Ежедневно в мире появляются гигабайты новых данных. Часть этих данных представляет собой текст, поэтому в настоящее время активно развивается такая область компьютерных наук, как обработка естественного языка. Одной из задач в этой области является анализ тональности текста. Эта задача подразумевает под собой автоматизированное выявление в текстах эмоционально окрашенной лексики и эмоциональной оценки авторов (мнений) по отношению к объектам, речь о которых идёт в тексте.

Большинство известных методов для анализа тональности, применяемых для русскоязычных текстов, классифицируют тексты по бинарной шкале (позитивно окрашенные\негативно окрашенные). Однако существуют методы для более глубокого анализа, которые извлекают эмоции из текстов. Например, Мостафа Аль Масум Шейх[1] описывает применение модели ОСС[2] для извлечения эмоций из английских текстов. Модель ОСС позволяет распознавать 22 типа эмоций, идентифицируя их путем оценки реакций на события, агентов или объектов, содержащихся в анализируемом тексте. Мостафе Аль Масум Шейху удалось достичь 80% точности распознавания эмоций с использованием ОСС модели. Однако, в научных публикациях не встречаются описания применения методов более глубокого анализа тональности с извлечением эмоций для русскоязычных текстов. Поэтому целью данного исследования является определение возможности использования ОСС модели для русскоязычных текстов.

ОСС модель описывает эмоции как иерархическую структуру, которая включает три класса, шесть групп и 22 типа эмоций. В этой модели эмоция представляет собой набор когнитивных переменных в определенном состоянии. Например, переменная *OtherPresump* может быть установлена как «желательная» или «нежелательная» при оценке события с точки зрения агента, имеющего отношение к оцениваемому событию. Так, для предложения «террорист сбежал из тюрьмы», значение *OtherPresump* устанавливается как «желательно», потому что отрицательный агент «террорист» связан с негативным действием «сбежать из тюрьмы».

Мостафа Аль Масум Шейх описывает правила обнаружения когнитивных переменных в тексте[1]. Эти правила включают грамматические, морфологические и статистические аспекты. Поэтому для ОСС модели необходимы грамматические и морфологические анализаторы, а также специальные словари. Для грамматического и морфологического анализа была использована программа SyntaxNet, а необходимые словари составлены из открытых словарей RuSentiLex[3] и LinisCrowd[4].

Для тестирования системы было подготовлено более 100 предложений с эмоциональной окраской. Эти предложения были оценены экспертами и результат был сопоставлен с работой ОСС модели. Сопоставление показало, что ОСС модель можно использовать для извлечения эмоций из русскоязычных текстов.

В дальнейших исследованиях планируется улучшить точность модели за счет расширения словарей и применить полученную модель для дополнения семантической сети.

Ссылки

1. Mostafa Al Masum Shaikh, Helmut Prendinger, and Mitsuru Ishizuka. Affective Information Processing. A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text 2009; 45–73.
2. A. Ortony, G.L. Clore, A. Collins. The cognitive structure of emotions, Cambridge University Press, 1988.
3. Loukachevitch, N. V., Levchik, A. V. Creating Russian Sentiment lexicon, Open Semantic Technologies for Intelligent Systems. BGUIR, 2016.
4. Koltsova O.Y., Alexeeva S.V., Kolcov S.N. An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. 2016. С. 277–287.

Автор _____ / _____ (Фамилия
И.О.)

(подпись)

Научный руководитель _____ / _____ (Фамилия
И.О.)

(подпись)

Заведующий кафедрой _____ / _____ (Фамилия
И.О.)

(подпись)

«_26_» _февраля_ 2019 года