

УДК 004.8

МОДЕЛЬ ЭМОЦИОНАЛЬНОГО АНАЛИЗА РУССКОЯЗЫЧНЫХ ТЕКСТОВ

Егорова Е.В.¹

Научный руководитель – Суриков А.Г.², к.т.н., научный сотрудник НЦКР

¹Университет ИТМО, Факультет цифровых трансформаций

²Университет ИТМО, Национальный центр когнитивных разработок

В настоящей работе представлена вероятностная модель эмоционального анализа русскоязычных текстов с учетом эмоциональных индикаторов. Данная модель осуществляет классификацию текстов по эмоциональной шкале Экмана, в которую входит 6 базовых эмоций (радость, печаль, страх, удивление, злость и отвращение), и определяет эмоциональную интенсивность текста.

Введение.

Социальные сети позволяют пользователю воссоздать свою личность в Интернете. Пользователи размещают у себя в профиле личные тексты, фотографии, ссылки, записи, видео и т. д. Личные тексты, которые размещают у себя пользователи, содержат большое количество информации не только об их смысловом содержании, но и о слабо формализуемом эмоциональном контексте их написания. В данной работе мы пытаемся извлечь эмоциональную подачу из текстов с помощью баесовской вероятностной модели, обученной на экспертно-размеченных текстах. Цель данной работы заключается в получении комплексного вывода об эмоциях, проявленных в текстах и их интенсивности, а также в получении промежуточных предикторов (эмбедингов) для улучшения качества вывода других моделей (например, моделей психотипирования пользователей по их цифровым следам).

Основная часть. При участии эксперта психолога был проведен анализ и сравнение, существующих моделей классификаций эмоций, в результате которых была выбрана эмоциональная модель Экмана, в которую входит 6 эмоций (радость, печаль, страх, удивление, злость и отвращение), как наиболее научно-обоснованная и часто используемая в научных исследованиях методика. Для данного исследования был произведен сбор и разметка данных, для удобства разметки был разработан чат-бот для разметки по эмоциональным классам. 5 экспертов разметили 6340 текстов. Сбор данных по минорным классам продолжается. Затем с помощью полученных данных, мы составили вероятностный словарь эмоциональных слов (эмоциональный лексикон), который, в дальнейшем может применяться для определения эмоциональных доминант и психотипирования пользователей. Было построено несколько моделей, основанных на разных подходах. Наилучшие результаты показала модель построенная на баесовских сетях доверия.

Выводы. Предложенная модель позволяет определять эмоции и их интенсивность в русскоязычных текстах с учетом эмоциональных индикаторов. Также на основании полученных результатов был сформирован вероятностный словарь эмоциональных слов (эмоциональный лексикон). Данные результаты будут основой для дальнейшего исследования и психотипирования пользователей.

Егорова Е. В.

Подпись

Суриков А.Г.

Подпись