

УДК 004

ИДЕНТИФИКАЦИЯ КЛЮЧЕВЫХ ЭМОЦИЙ ЧЕЛОВЕКА В КОНТЕКСТЕ ЗАДАЧИ ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ ТЕКСТА

АГНАЕВ АЛЕКСАНДР ЮРЬЕВИЧстудент
Национальный исследовательский университет ИТМО*Научный руководитель: Штенников Дмитрий Геннадьевич*
доцент факультета ПИиКТ, к.т.н.
НИУ ИТМО

Аннотация: задача определения тональности текста все чаще встречается в различных областях жизни человека, например в маркетинге, психологии, клиентском сервисе и в других деятельности. Однако, интерпретация эмоционального окраса имеет множество проблем, напрямую влияющих на точность определения. Причиной для этого служат различные неоднозначности речи, различие чувств по силе проявления, неявности выражения или проявления нескольких эмоций одновременно.

Для облегчения анализа тональности текста, необходима составленная палитра эмоций человека, элементы которой будут достаточно сильно отличаться друг от друга, чтобы избежать спорных ситуаций, а также иметь явные слова-маркеры, обозначающие конкретную эмоцию, для возможности более точного распознавания.

Ключевые слова: эмоция, тональность, классификация, датасет, базовые эмоции.

IDENTIFICATION OF KEY HUMAN EMOTIONS IN THE CONTEXT OF THE TASK OF DETERMINING THE TONE OF A TEXT

Agnaev Alexander Yuryevich*Scientific adviser: Shtennikov Dmitry Gennadievich*

Abstract: The task of determining the tone of a text is increasingly common in various areas of human life, for example, in marketing, psychology, customer service and other activities. However, the interpretation of emotional coloration has many problems directly affecting the accuracy of the determination. The reasons for this are various ambiguities of speech, the difference of feelings by the strength of manifestation, implicit expression or manifestation of several emotions at the same time.

To facilitate the analysis of the tone of the text, it is necessary to compose a palette of human emotions, the elements of which will be sufficiently different from each other to avoid controversial situations, as well as to have explicit marker words denoting a particular emotion, for the possibility of more accurate recognition.

Key words: emotion, tone, classification, dataset, basic emotions.

Человек проявляет свои эмоции ежедневно, в каждом разговоре. Они играют очень большую роль в поведении, мышлении и том, как люди взаимодействуют с окружающими. Именно поэтому эмоции человека являются распространенной темой среди психологических исследований и исследований искусственного интеллекта.

Одной из самых ранних работ на эту тему является книга Уильяма Джеймса “Принципы психологии”, вышедшая в 1890 году. Теория эмоций Джеймса на тот момент была крайне новаторской, он утверждал, что эмоции являются не более чем физиологическими реакциями на внешние стимулы, а также был уверен, что они предшествуют когнитивной оценке ситуации. Помимо этого, им были выдвинуты четыре основные эмоции, такие как радость, печаль, гнев и страх [1, с. 449]. Данный набор является самыми базовыми ощущениями, они покрывают большой спектр человеческих чувств, а также являются обособленными и отличающимися друг от друга. Радость и печаль являются противоположными, с явными текстовыми обозначениями эмоции, что крайне благоприятно будет влиять на их определение. Гнев и страх в свою очередь являются крайне специфичными, и по своей природе более яркими относительно других эмоций. Соответственно, их определение является более простой задачей, и увеличит точность при анализе тональности. Данные четыре эмоции в последствии будут использоваться и в других теориях, соответственно их можно принять за основу палитры для решения задачи.

Несмотря на то, что выявленные Джеймсом основные эмоции были бы крайне удачным выбором для определения тональности текста, они упускают многие другие чувства, которые может испытывать человек. В конце XX века, профессор Пол Экман, американский психолог, выдвинул свою теорию эмоций, которая в том числе ссылается на теорию Джеймса, и дополняет ее. Он утверждает, что базовые эмоции являются универсальными среди людей любых культур и народов. В основе его теории лежат 6 различных эмоций, таких как основные злость, радость, страх, печаль, а также дополнительные отвращение и удивление [2, с. 22]. Однако, помимо основной четверки, упомянутой еще Уильямом Джеймсом, отвращение и удивление являются гораздо более комплексными эмоциями. Крайне сложно выделить слова или словосочетания, напрямую передающее отвращение. Удивление, в свою очередь, является проблематичным из-за схожести с радостью. Также, в контексте оно может быть похоже на испуг, т.к. в основе обеих эмоций лежат разрушенные ожидания или внезапность, что значительно затрудняет классификацию предложений. Данные эмоции будет проблематично определять используя исключительно текст, так основной способ их выражения напрямую связан с мимикой.

Рассмотрим другую работу Экмана [3, с. 55], в которой он выделяет список из 15 основных эмоций. Так же, как и с презрением и удивлением, многие чувства из списка с трудом поддаются идентификации при помощи текста, например гордость или смущение. Они либо частично пересекаются с основными эмоциями, либо имеют большие сложности в выявлении напрямую передающих чувства слов (слов-идентификаторов). Но также, в списке присутствуют такие эмоции как веселье и стыд, которые соответствуют всем требованиям, необходимым для использования в решении задачи. Они проявляются достаточно ярко, для того чтобы их было удобно классифицировать, а также сильно отличаются от других эмоций, которые мы уже добавили в палитру.

Таблица 1

Выделенные эмоции и примеры слов-идентификаторов

| Эмоция | Радость | Грусть | Злость | Страх | Веселье | Стыд | Отвращение | Удивление |
|----------------------------------|--|---|--|---|--|--|--|---|
| Слова для идентификации в тексте | Круто, классно, радость, приятно, нравится, лучший, весело, невероятно, удивительно, вау | Печально, грустно, тяжесть, расстроиться, жалость, сложность, гадость | Злиться, ругательства, раздражать, ярость, рассердиться, возмутиться, ненавидеть | Бояться, страшно, боязнь, испугаться, жутко, стремно, тревожиться, внезапно, неожиданно | Смешно, угар, весело, смеяться, забавно, ржать | Прости, виноват, жалкость, неловкость, стесняться, стыдиться | Гадость, ненавидеть, раздражать, противность | Невероятность, удивительно, внезапно, неожиданность |
| Количество | 8 | 7 | 7 | 8 | 6 | 6 | 4 | 4 |
| Количество совпадений | 3 | 1 | 2 | 2 | 1 | 1 | 3 | 4 |

Рассмотрев все упомянутые эмоции и их слова-идентификаторы (табл. 1), можно прийти к выводу, что эмоции отвращения и удивления не подходят для решения задачи, из-за высокого процента совпадений слов идентификаторов и другими эмоциями. Веселье и стыд были добавлена в палитру, т.к. они легче всего отличаются от выбранных нами ранее чувств.

Таким образом, после изучения трудов Экмана и Джеймса и сравнительного анализа полученных эмоций, мы можем прийти к выводу, что необходимая нам палитра эмоций состоит из злости, радости, страха, печали, веселья и стыда. Однако, учитывая специфику задачи, а именно необходимость анализа текстовых данных, нужно принимать во внимание то, что часть текста не несет в себе какой-либо эмоциональной окраски, и является, например, сухой констатацией какого-либо факта. Чтобы повысить точность определения тональности текста, необходимо добавить в палитру нулевую эмоциональную окраску. Назовем ее нейтральностью.

В результате исследования была проведена идентификация ключевых эмоций человека для решения задачи определения тональности текста, в результате чего получили палитру из 7 эмоциональных состояний: нейтральность, радость, печаль, злость, страх, веселье, стыд.

Для того, чтобы выбранная нами палитра эмоций могла быть распознана во время обучения корректно, необходимо использовать специфичные данные для обучения. Датасет был составлен из коротких, эмоционально окрашенных текстов, например сообщений в мессенджерах, социальных сетях или комментариях. Из-за отсутствия наборов данных, содержащих подходящие данные на изначально русском языке, был выбран датасет переведенных сообщений с платформы Reddit ru_go_emotions из открытой библиотеки Hugging Face. Так как данный набор данных является переводом go_emotions, и приведенные в нем оценки эмоций были сделаны относительно английского текста, была проведена переоценка эмоций. Каждому тексту был присвоен номер от 0 до 6, где 0 – нейтральность, 1 – радость, 2 – печаль, 3 – злость, 4 – страх, 5 – веселье, 6 – стыд. Номера соответствуют основному эмоциональному окрасу сообщения. Если в тексте сообщения находилась серьезная ошибка, или английский текст, производился перевод или исправление. В случае спорной ситуации, когда не было возможности определить точно, к какой эмоции относится тот или иной текст, он удалялся из датасета и заменялся другим сообщением.

Список источников

1. James, W. (1890). The principles of psychology, Vol. 2. Henry Holt and Company.
2. Ekman, P., Friesen, W. V. (2003). Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues. United States: Malor Books.
3. Ekman, Paul. "Basic emotions." Handbook of cognition and emotion 98.45-60 (1999): 16.