Научная статья
Original article
УДК 004.934

РАСПОЗНАВАНИЕ ЭМОЦИЙ В ТЕКСТОВОМ СООБЩЕНИИ

EMOTION RECOGNITION IN TEXT MESSAGE

Дворников Сергей Викторович, магистрант, Донской государственный технический университет, г. Ростов-на-Дону, sergey.dv520@gmail.com

Dvornikov Sergey Viktorovich, Master's student, Don State Technical University, Rostov-on-Don, sergey.dv520@gmail.com

Аннотация. Данная статья посвящена теме распознавания эмоций в различных текстовых сообщениях. Описаны сложности распознавания эмоций, представлены различные формулировки задач, решаемых в рамках данной проблемы. Рассмотрены основные подходы, используемые для извлечения эмоций из представленного текста и приведены модели машинного обучения.

Abstract. This article focuses on the topic of recognizing emotions in various text messages. The difficulties of recognizing emotions are described, various formulations of the tasks solved in the framework of this problem. The main approaches used to extract emotions from the presented text are considered and machine learning models are presented.

Ключевые слова: распознавание эмоций, анализ тональности, машинное обучение, описание эмоций, классификация.

Keywords: emotion recognition, sentiment analysis, machine learning, emotion description, classification.

Проблема распознавания эмоций в текстах или сообщениях является чрезвычайно сложной, но актуальной. Эмоциональные расчеты используются в робототехнике, транспорте, игровой индустрии, образовании, маркетинговых исследованиях, поисковых системах, человеко-машинных интерфейсах и др. Исследования в этой области позволяют обучать искусственный интеллект распознаванию человеческих эмоций и настраивать работу различных информационных систем в зависимости от состояния человека. Сложность определения эмоции в сообщении обосновывается следующими причинами [1]:

1) Выражение эмоций сильно зависит от контекста.
2) Частота слов важнее их расположения.
3) Наличие различных стилистических приемов.

Тема распознавания эмоций в сообщениях довольно обширна и в ее рамках выделяют несколько задач.

Приведем основные типы задач:

1) Анализ тональности текста – класс методов контент-анализа в компьютерной лингвистике, предназначенный для автоматизированного выявления в текстах эмоционально окрашенной лексики и эмоциональной оценки авторов (мнений) по отношению к объектам, речь о которых идёт в тексте [2].

Анализ тональности текста может найти применение в следующих областях:

– способен помочь научить компьютер воспринимать естественный язык на уровне, приближенном к человеческому. До сих пор машина понимала тексты на абстрактном уровне – в основном, через лексемы (слова), которые для нее обладали формой (набор букв) и содержанием (значение).

Данная концепция предлагает ввести еще одну функцию – лексическую тональность текста;

– способен значительно повысить качество машинного перевода. При переводе не обойтись без первичного анализа текста и отдельных слов – в том числе, анализа тональности;

– приблизить образ мышления компьютера к человеческому, как некое мнение автора.

2) Анализ субъективности – определить, является ли текст субъективным или объективным, т.е. выражены ли в тексте эмоции

3) Классификация по эмоциям – определить, какая конкретно эмоция выражена в тексте. Исследования Пола Экмана показывают, что используются 6 базовых эмоций [3] – радость, злость, отвращение, страх, грусть и удивление (таблица 1).

4) Извлечение мнений – определить, кто по отношению к кому или чему выразил в тексте мнение, и какова тональность этого мнения.

Таблица 1 – Примеры предложений с эмоциями.

Пример эмоции	**Эмоция**
Я очень рада за твои достижения!	Радость
Я очень рассержен на тебя!	Злость
Как ты можешь это есть?	Отвращение
Эта ситуация вызывает у меня плохое предчувствие!	Страх
Мне очень жаль…	Грусть
Как до сих пор это работает?	Удивление

Основываясь на инструкциях Экмана по основным человеческим эмоциям, можно создать краткую аннотацию эмоций. Текст, который анализируется, может быть любой – сообщение, комментарий, отзыв и т.д.

– радость: относят к состояниям, вызванным ощущением удовольствия. Эти состояния варьируются от удовольствия от помощи другим, теплого приподнятого чувства, которое люди испытывают, когда

видят доброту и сострадание, переживания легкости и удовлетворенности или даже наслаждения несчастьями другого человека до радостной гордости за свои достижения или переживание чего-то очень красивого и удивительного;

– злость: состояния, вызванные чувством неуспеха в нашем прогрессе. Содержит как раздражение, так и ярость и варьируется от разочарования, которое является реакцией на неоднократные неудачи преодолеть препятствие до гнева, вызванного сильным неудобством, от споров до горечи - гнева после несправедливого обращения и мстительности;

– отвращение: для сообщений, которые выражают как неприязнь, так и отвращение. Они варьируется от побуждения избегать чего-то отвратительного или отвращения от реакции на дурной вкус, запах, вещь или идею;

– страх: свидетельствует о тревоге и ужасе. Состояния варьируются от трепета - ожидания возможной опасности, нервозности, до отчаяния, реакции на неспособность уменьшить опасность, паники и ужаса. Смесь страха, отвращения и шока;

– грусть: содержит одновременно разочарование и отчаяние. Состояния варьируются от уныния, безумия, беспомощности, безнадежности до сильных страданий, чувства и печали, часто вызванных утратой или печалью и болью;

– удивление: комментарии, которые выражают чувства, вызванные неожиданными событиями, во что-то, во что трудно поверить и которые могут вас шокировать. Это самая короткая эмоция из всех, она длится всего несколько секунд. Это может быть страх, гнев, облегчение или ничем, в зависимости от события, которое нас удивляет;

– другое: для текстов, которые не показывают ни одной из вышеперечисленных эмоций или не содержат эмоций.

Перед извлечением эмоций из текста необходимо его отфильтровать. Сообщения в большинстве случаев зашумлены (например, лишние или пропущенные символы или буквы, ошибки правописания, смайлики,

повторения знаков, хештеги и др). После «очищения» получаются отфильтрованные чистые данные, с которыми удобно работать.

Подходы к извлечению эмоций из текста.

Выделяют несколько подходов [3]:

1) Подход, основанный на ключевых словах**.** Наиболее интуитивный и сильный подход. Основывается в нахождении шаблонов близких к эмоциональным ключевым словам и сопоставлении их.

2) Подход, основанный на лексике. Классифицирует текст используя имеющийся лексикон (база знаний текстов, помеченных в соответствии с эмоциями) для входных данных.

3) Машинное обучение. Для распознавания эмоций используются обучение как с учителем, так и без, в которых модель направлена на обучение и тестирование классификатора, при этом данные делятся на обучающую и проверочную выборки.

4) Смешанный подход. Комбинирует два или три метода для достижения наилучшего выигрыша среди множества алгоритмов и наивысшего уровня точности.

С первыми двумя методами все более-менее понятно. Поподробнее остановимся на машинном обучении.

Для обучения алгоритмов машинного обучения необходима выборка.

Она нужна для тренировки модели Machine Learning, чтобы обучить систему и затем использовать ее для решения реальных задач [5].

Выборку нужно разбить на 7 классов эмоций: 1 – радость, 2 – злость, 3 – отвращение, 4 – страх, 5 – грусть, 6 – удивление, 7 – нейтральные.

Можно использовать модели машинного обучения такие как:

1) Машина опорных векторов (SVM)

2) Случайный лес

И две модели глубокого обучения:

1) Сверточная нейронная сеть (CNN)

2) Рекуррентная нейронная сеть (LSTM – «Долгосрочная краткосрочная память»).

Метод опорных векторов – это метод машинного обучения, предназначенный для решения задач классификации и регрессии

Исследователи из OpenAI применяли для распознавания эмоций LSTM и вид данной нейронной сети хорошо зарекомендовал себя в распознавании написанного текста и человеческой речи. Ей принадлежит рекорд минимума ошибок при распознавании речи – 17,7% [6].

Важная часть машинного обучения – это классификация. Если нужно знать, к какому классу принадлежит значение. Задачу классификации хорошо решает случайный лес. Фундаментальная концепция в основе случайного леса проста, но сильна – это знание большинства [7].

Для решения задачи автоматического распознавания эмоций в рассматриваемой системе используется сверточная нейронная сеть. Это особая архитектура нейронных сетей, основным предназначением которой является эффективное распознавание образов [8].

Проблема автоматического распознавания эмоций в текстах является в настоящее время весьма важной и актуальной. Для ее решения было разработано множество методов и алгоритмов. Все они обладают своими достоинствами и недостатками, и одно из перспективных направлений исследований заключается в комбинировании методов разных подходов.

Литература

1. Обзор подходов для автоматического распознавания эмоций в текстах [электронный ресурс] // https://cyberleninka.ru/article/n/obzor-podhodov-dlya-avtomaticheskogo-raspoznavaniya-emotsiy-v-tekstah (дата обращения 16.11.2021)
2. Анализ тональности текста: концепция, методы, области применения [электронный ресурс] // http://datareview.info/article/analiz-tonalnosti-teksta-kontseptsiya-metodyi-oblasti-primeneniya/ (дата обращения 16.11.2021)

3. Википедия site:star-wiki.ru – Распознавание эмоций - Emotion recognition [электронный ресурс] // https://star-wiki.ru/wiki/Emotion_recognition#Emotion_recognition_in_text (дата обращения 17.11.2021)
4. Использование современных алгоритмов машинного обучения для задачи распознавания эмоций [электронный ресурс] // https://cyberleninka.ru/article/n/ispolzovanie-sovremennyh-algoritmov-mashinnogo-obucheniya-dlya-zadachi-raspoznavaniya-emotsiy/viewer (дата обращения 18.11.2021)
5. Отберем то, что нужно Data Mining: как сформировать датасет для машинного обучения [электронный ресурс] // https://www.bigdataschool.ru/blog/dataset-data-preparation.html (дата обращения 18.11.2021)
6. Искусственный интеллект научили лучше распознавать эмоции в тексте [электронный ресурс] // https://aboutdata.ru/2017/04/08/ai-sentiment-recognition/ (дата обращения 19.11.2021)
7. Как работает случайный лес? [электронный ресурс] // https://nuancesprog.ru/p/6160/ (дата обращения 20.11.2021)
8. Использование нейросетевых технологий в задаче автоматического распознавания эмоций / С. О. Целикова, Я. П. Горожанкин, А. О. Иванов [и др.]. — Текст: непосредственный // Молодой ученый. — 2019. — № 26 (264). — С. 59-61. — URL: https://moluch.ru/archive/264/61173/ (дата обращения: 20.11.2021).

Literature

1. Review of approaches for automatic recognition of emotions in texts [electronic resource] // https://cyberleninka.ru/article/n/obzor-podhodov-dlya-avtomaticheskogo-raspoznavaniya-emotsiy-v-tekstah (date of access 16.11.2021)

2. Analysis of the sentiment of the text: concept, methods, areas of application [electronic resource] // http://datareview.info/article/analiz-tonalnosti-teksta-kontseptsiya-metodyi-oblasti-primeneniya/ (date of treatment 11/16/2021)
3. Wikipedia site: star-wiki.ru - Emotion recognition [electronic resource] // https://star-wiki.ru/wiki/Emotion_recognition#Emotion_recognition_in_text (accessed 17.11.2021)
4. Using modern machine learning algorithms for the emotion recognition problem [electronic resource] // https://cyberleninka.ru/article/n/ispolzovanie-sovremennyh-algoritmov-mashinnogo-obucheniya-dlya-zadachi-raspoznavaniya-emotsiy/viewer (date reference 18.11.2021)
5. Let's select what you need Data Mining: how to form a dataset for machine learning [electronic resource] // https://www.bigdataschool.ru/blog/dataset-data-preparation.html (date of treatment 11/18/2021)
6. Artificial intelligence was taught to better recognize emotions in the text [electronic resource] // https://aboutdata.ru/2017/04/08/ai-sentiment-recognition/ (date of treatment 11/19/2021)
7. How does the random forest work? [electronic resource] // https://nuancesprog.ru/p/6160/ (date of treatment 11/20/2021)
8. The use of neural network technologies in the problem of automatic recognition of emotions / SO Tselikova, Ya. P. Gorozhankin, AO Ivanov [and others]. - Text: direct // Young scientist. - 2019. - No. 26 (264). - S. 59-61. - URL: https://moluch.ru/archive/264/61173/ (date of access: 20.11.2021).

© Дворников С.В., 2021 Научно-образовательный журнал для студентов и преподавателей «StudNet» №11/2021.

Для цитирования: Дворников С.В., РАСПОЗНАВАНИЕ ЭМОЦИЙ В ТЕКСТОВОМ СООБЩЕНИИ // Научно-образовательный журнал для студентов и преподавателей «StudNet» №11/2021.