

АВТОМАТИЧЕСКИЕ МЕТОДЫ БИМОДАЛЬНОГО РАСПОЗНАВАНИЯ ЭМОЦИЙ В РЕЧИ

Шевченко Олег Владимирович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: shevchenko.cmc@gmail.com

Научный руководитель — Лукашевич Наталья Валентиновна

В последние годы задача автоматического распознавания эмоций все больше вызывает интерес исследователей в области искусственного интеллекта. Наибольшая возможность применения подобных технологий есть в таких областях как медицина, распознавание состояния водителя, маркетинговые исследования, системы видеоаналитики, носимые устройства и др.

Методы автоматического распознавания эмоций обычно основаны на одной или нескольких модальностях. Модальность - это форма выражения некоего действия, такого как написанный текст, устная речь или выражение лица человека. Таким образом методы распознавания эмоций могут делиться на одномодальные, которые распознают эмоции на каком-то одном конкретном входе, например, устной речи, и на мультимодальные, которые распознают эмоции из нескольких входных каналов, к примеру, по видео, беря в расчет голос и мимику говорящего. В данной работе была рассмотрена бимодальная модель распознавания эмоций, которая основана на устной и письменной речи.

Рассматривалось следующее множество классов эмоций: грусть, счастье, страх, удивление, гнев, отвращение, а также нейтральное состояние. В качестве размеченного набора примеров была использована база данных RAMAS [2]. В результате было получено 10835 аудио фрагментов продолжительностью от 3 до 10 секунд. Фрагменты представлены в виде диалога двух людей.

Из голосовых сообщений извлекались различные признаки такие, как список мел-кепстральных коэффициентов, энергия, подача и др. Так как в используемом наборе данных не было текстовой информации, то текст распознавался из аудио примеров при помощи сторонних методов по распознаванию речи [3]. Распознанный текст переводился в векторное представление tf-idf [1]. Объединенный вектор признаков из двух каналов подавался на вход общему мультиклассовому классификатору. В данной работе были рассмотрены такие методы классификации, как деревья решений, градиентный бустинг,

метод опорных векторов, наивный байес, линейная регрессия, многослойный перцептрон и рекуррентная нейронная сеть.

Для анализа результатов использовались такие метрики, как точность, полнота, F-мера и др. Результаты работы, представленные на Рис. 1, показывают несовершенство предложенных подходов и дают возможность для их улучшения путем построения более сложных архитектур нейронных сетей и рассмотрения большего набора признаков.

Иллюстрации

Models	Accuracy	F1	Precision	Recall
RF	51.0	51.2	51.2	51.3
XGB	53.6	52.8	52.9	52.8
SVM	33.7	15.2	17.4	21.5
MNB	31.3	9.1	19.6	17.2
LR	33.4	14.9	17.8	20.9
MLP	41.0	36.5	42.2	35.9
LSTM	53.6	51.3	53.2	49.6

Рис. 1: Результаты метрик на различных алгоритмах.

Литература

1. Manning C. D., Raghavan P., Schütze H. Introduction to information retrieval. Cambridge: Cambridge University Press, 2008.
2. Perepelkina O., Kazimirova E, Konstantinova M. 2018. RAMAS: Russian Multimodal Corpus of Dyadic Interaction for studying emotion recognition. PeerJ Preprints 6:e26688v1.
3. Страница облачного сервиса по распознаванию речи Yandex.SpeechKit:
<https://cloud.yandex.ru/docs/speechkit/>