ТЕХНІЧНІ НАУКИ

Семенюк В.В.
магистр,
Донецкий национальный университет

ЭФФЕКТИВНОСТЬ ПРИМЕНЕНИЯ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИЙ ЧЕЛОВЕКА ПО ЕГО РЕЧИ

Введение. Эмоции и речь тесно взаимосвязаны и играют огромную роль в общении. В связи с этим, автоматическая и объективная диагностика эмоционального состояния человека по его речи представляет большой практический интерес. Возможность распознавания эмоций в речи важна как для исследования самой речи и эмоций, так и для улучшения качества обслуживания клиентов, например, в телекоммуникационной сфере. Идентификация эмоционального состояния крайне востребована в индустрии развлечений, обучении, медицине и других сферах.

Цель работы состоит в исследовании эффективности использования сверточных нейронных сетей и алгоритма вычисления значимых характеристик речевого сигнала методом звуковых отпечатков в задаче автоматического распознавания эмоций в речи.

Описание технологии распознавания эмоций по голосу и полученных результатов. Для оценки эмоционального состояния человека эмоции были разбиты на 3 общие группы: позитивные, негативные и нейтральные эмоции. Затем, для более точной классификации, эмоции делились на восемь классов: агрессия, отвращение, страх, счастье, нейтральное состояние, грусть, подавленность и удивление.

Для вычисления значимых признаков распознавания был использован метод звуковых отпечатков, который заключается в том, что обучающая выборка состоит из спектрограмм аудиозаписей для каждой эмоции [1].

Для распознавания как общих групп, так и восьми классов эмоционального состояния человека были созданы две нейронные сети. Используемый тип нейронных сетей – свёрточная [2]. Архитектуры обоих нейросетей идентичны за исключением количества выходных классов. На рисунке 1 изображена используемая архитектура сети.

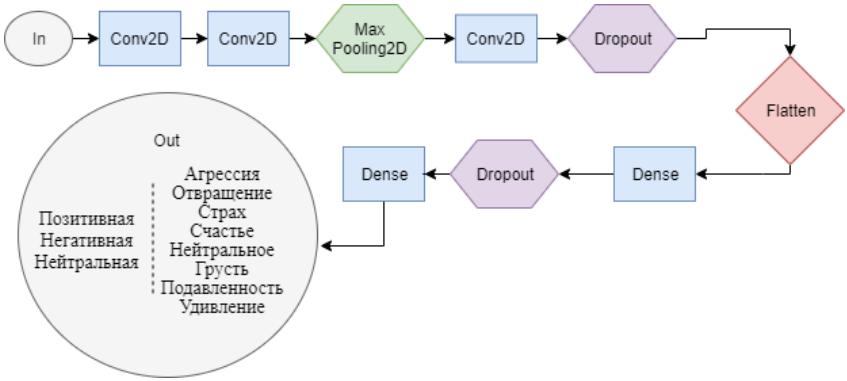

Рис. 1. Архитектура нейросети для распознавания эмоций человека по голосу

Для обучения нейронных сетей использовалась выборка, состоящая из 4500 аудиофайлов, каждый из которых содержал фрагмент эмоционально окрашенной слитной фразы, частота дискретизации 48kHz, уровень квантования – 16 бит, длина фрейма – 512 отчётов. Использовались записи двадцати четырёх профессиональных актёров (12 женщин и 12 мужчин). На рисунке 2 представлен пример изображения из обучающей выборки.

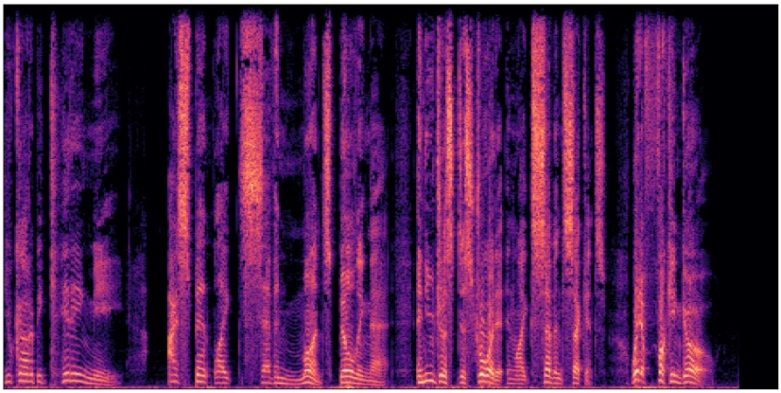

Рис. 2. Изображение из обучающей выборки

Объем тестовой выборки составил 956 записей эмоционально окрашенных слитных фраз. На рисунке 3 показаны результаты

правильного распознавания (в процентах) для группы из трёх классов (слева) и для группы из восьми классов эмоций (справа).

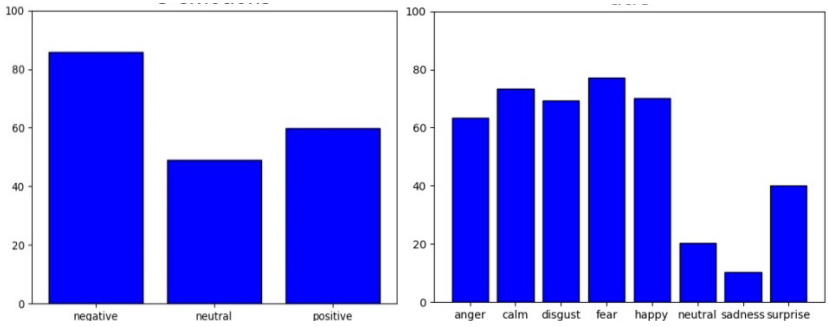

Рис. 3. Результаты правильного распознавания для 3 и 8 классов эмоций

Из полученных результатов видно, что наибольшую точность распознавания имеют негативные эмоции, наихудшие показатели – у нейтральных эмоций.

Заключение. Применение подхода к классификации эмоционального состояния человека по его голосу с использованием машинного обучения и метода звуковых отпечатков показало хороший результат (80%) эффективности распознавания только для негативных эмоций, для эмоций остальных групп эффективность распознавания ниже. Предположительно, для улучшения точности распознавания можно применить дополнительные характеристики фреймов сигнала, полученные алгоритмом MFCC. Алгоритм MFCC учитывает волновую природу звука и психофизическое восприятие звука человеком, устойчив к изменению тембра голоса, громкости и скорости произношения [3].

Список использованных источников:

1. Peter Grosche, Meinard Müller, and Joan Serrà, «Audio Content-based Music Retrieval», Multimodal Music Processing, Meinard Müller, Masataka Goto, and Markus Schedl, Eds, vol. 3 of Dagstuhl Follow-Ups, chapter 9, pp. 157–174. Dagstuhl Publishing, Wadern, Germany, April 2012.

2. Dan C. Ciresan, Ueli Meier, Jonathan Masci, et al. Flexible, High Performance Convolutional Neural Networks for Image Classification. PROCEEDINGS OF THE TWENTY-SECOND IN TERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2011.

3. Оппенгейм А.В. Цифровая обработка сигналов / А.В. Оппенгейм, Р. Шафер. – Москва: Техносфера, 2012. – 1048 с.