

УДК 004.021

Егорчев Антон Александрович, Чикрин Дмитрий Евгеньевич,
Пашин Дмитрий Михайлович, Сарамбаев Никита Андреевич,
Фахрутдинов Адель Фердинандович

АНАЛИЗ СИСТЕМ ОПРЕДЕЛЕНИЯ И КЛАССИФИКАЦИИ ЭМОЦИЙ ЧЕЛОВЕКА ПО ДАННЫМ ЗВУКОВОГО ПОТОКА

Возможность быстрой определения эмоционального состояния человека для выявления негативного настроения позволяет снизить материальные риски и даже сохранить жизнь людей, что является актуальной задачей на сегодняшний день. Целью работы является разработка алгоритма классификации эмоций по аудиоданным смартфона пользователя. В статье представлено решение, пригодное для встраивания в систему мониторинга состояния здоровья человека, которое позволяет определять эмоции человека по акустическому сигналу с микрофона пользователя. Решение базируется на использовании сверточных нейронных сетей. Решение показало 19.5% и 20.1% ошибок первого и второго рода соответственно для определения негативных эмоций.

Система неинвазивного мониторинга, машинное обучение, биомедицинский мониторинг, смартфонная сенсорика, анализ акустического сигнала, распознавание эмоций.

Egorchev Anton Alexandrovich, Chikrin Dmitry Evgen'evich,
Pashin Dmitry Mikhailovich, Sarambaev Nikita Andreevich
Fakhrutdinov Adel Ferdinandovich

EMOTION DETECTION AND CLASSIFICATION SYSTEM BASED ON SOUND FLOW DATA

The ability to quickly determine the emotional state of an employee to identify a negative allows us to reduce material risks and even save people's lives, which is an important task for today. The aim of the work is to develop an algorithm for classifying emotions based on the audio data of the user's smartphone. The paper presents a solution suitable for embedding into a human health monitoring system, which allows you to determine a person's emotions by an acoustic signal from the user's microphone. The solution is based on the use of convolutional neural networks. The solution showed 19.5% and 20.1% of errors of the first and second kind, respectively, for determining negative emotions.

Noninvasive monitoring system, machine learning, biomedical monitoring, smartphone sensors, acoustic signal analysis, emotion recognition.

Введение

В современном развивающемся мире имеется огромное количество рабочих должностей для разного уровня специалистов. Очевидно, что стресс оказывает негативное влияние на деятельность работника, а в конечном счете на производительность компании. В бизнесе и менеджменте эмоциональный интеллект играет важную роль в формировании лидерства и взаимоотношениях с коллегами и подчиненными. В образовании эмоции позволяют создавать позитивную образовательную среду и эмоционально поддерживать учащихся. В творческих профессиях эмоции напрямую влияют на производимый продукт. Более того негативные эмоции могут быть предвестниками конфликтов. Следовательно, существует необходимость в разработке автоматизированной системы, которая позволит своевременно выявлять людей с негативными эмоциями, потенциальным психоэмоциональным стрессом, для снижения негативного влияния таких людей на рабочую среду, а также оказания своевременной помощи, например, для предотвращения суицидов. В данной работе предложено решение на основе анализа алгоритмов определения эмоций из категории алгоритмов искусственного интеллекта, который служит для распознавания одной из трех эмоций: положительную, нейтральную или негативную в речевом сигнале. Данное решение предназначено для применения в рамках большой системы мониторинга состояния здоровья человека на предприятиях.

Основная часть

В задачах по распознаванию речи если не рассматривать модели основанные на трансформерах и модели seq-to-seq, используются рекуррентные нейронные сети или рекуррентные модели с долгой краткосрочной памятью (LSTM). В качестве признаков, извлекаемых из речевого сигнала, используются мел-кепстральные коэффициенты или статистические признаки речевого сигнала. На данных признаках как раз и обучают модели на основе нейронных сетей. Задача распознавания эмоций отлична от задачи распознавания речи тем, что при распознавании эмоций возможно не обязательно знать контекст, при котором эти эмоции появляются, что дает возможности рассмотреть сверточные нейронные сети для детекции эмоциональной окраски и ее распознавания.

Предлагается следующая последовательность действий для построения алгоритма распознавания эмоций:

- Построение табличной структуры данных: строится таблица данных со значениями эмоций для каждой аудиозаписи.
- Предварительная обработка: используя инструменты OpenSMILE[1], извлекаются статистические признаки из аудио-сигнала
- Обучение модели: используя CNN модель и построенную таблицу, модель обучается на распознавание эмоций.
- Тестирование модели: используя тестовые данные, модель тестируется на способность распознавания эмоций.
- Использование модели: обученная модель может использоваться для распознавания эмоций в новых аудио-сигналах.

При построении системы по распознаванию эмоций были рассмотрены следующие решения: решение Siddhant Mulajkar [2] (мел-кепстральные коэффициенты + CNN), решение на платформе Kaggle от Eu Jin Lok [3]; были рассмотрены следующие датасеты: SAVEE [5], RAVDESS [6], CREMA-D [7].

В качестве признаков для обучения сверточной нейронной сети были рассмотрены мел-кепстральные коэффициенты и статистические признаки из библиотеки openSMILE [1].

Временной отрезок в аудио-сигнале для проведения детекции эмоций 2.5 сек. Частота дискретизации аудио-сигнала 44100 Гц.

Структурная схема построенного алгоритма представлена на рис.1.

Результаты испытаний по определению и классификации эмоций представлен в таблице 1. Эмоция определяется как нейтральная, если в отрезке аудио-сигнала не существует эмоциональной окраски, иначе распознается либо положительная, либо отрицательная эмоция. Эмоциональная составляющая отрезка, большего чем 2.5 сек определяется вычисления моды по результатам проанализированных отрезков в 2.5 сек.

Таблица 1

Результаты испытания классификации эмоций

Тип эмоции	Ошибки 1 рода	Ошибки 2 рода
Негативная	19.5 %	20.1 %
Положительная	22.4 %	31.5 %
Нейтральная	40.6 %	53.3 %

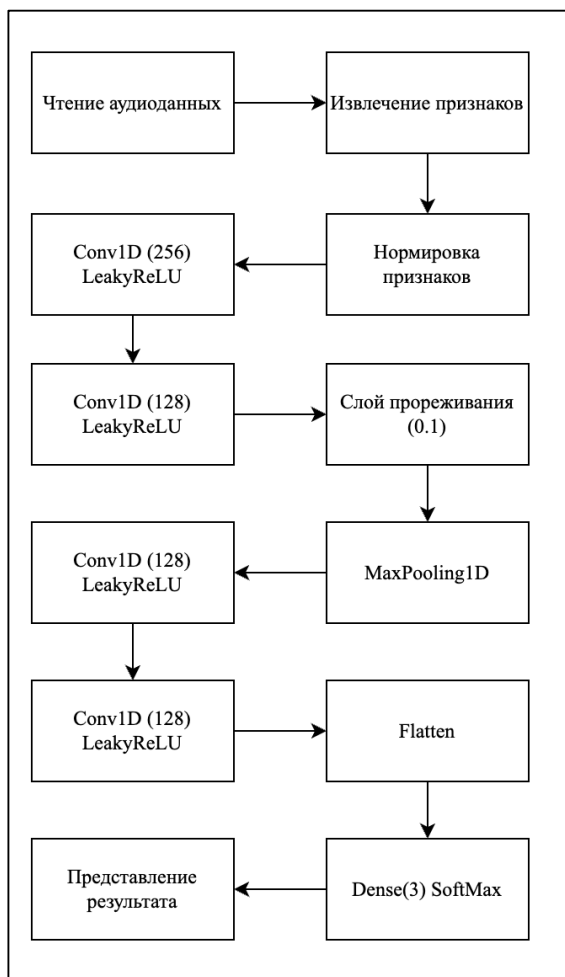


Рис 1. Структурная схема алгоритма определения и классификации эмоций

Выводы

В результате проведенных испытаний можно сделать вывод о том, что представленная модель позволяет практически с 20% ошибкой определять негативные эмоции. Что в принципе является более приоритетным классом для определения в рамках рассматриваемого применения решения. Однако является непригодной для определения других типов эмоций.

Основными причинами таких результатов может являться небогатая обучающая выборка. Потенциально можно улучшить результаты, расширив количество рассматриваемых признаков, проведя балансировку обучающей выборки, а также увеличив размер ее размер. Возможно решение можно улучшить, рассмотрев использование LSTM на эмбедингах сверточной нейронной сети, или использовав комбинацию CNN+LSTM.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Opensmile 2.4.2 // PyPI - Указатель пакетов Python'a URL: <https://pypi.org/project/opensmile/> (дата обращения: 12.12.2022).
2. Sentiment-Predictor-for-Stress-Detection-using-Voice. — Текст : электронный // GitHub : [сайт]. — URL: <https://github.com/sidmulajkar/sentiment-predictor-for-stress-detection> (дата обращения: 10.10.2022).
3. Audio Emotion | Part 1 - Explore data. — Текст : электронный // Kaggle : [сайт]. — URL: <https://www.kaggle.com/code/ejlok1/audio-emotion-part-1-explore-data> (дата обращения: 10.10.2022).
4. Теория и практика машинного обучения: учебное пособие / В. В. Воронина, А. В. Михеев, Н
5. and Signal Processing URL: <http://kahlan.eps.surrey.ac.uk/>. Г. Ярушкина, К. В. Святков. — Ульяновск: УлГТУ, 2017. — 290 с.
6. Surrey Audio-Visual Expressed Emotion (SAVEE) Database // The Centre for Vision, Speech savee/ (дата обращения: 29.11.2022).
7. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English // PLOS URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0196391> (дата обращения: 28.11.2022).
8. CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset // The National Center for Biotechnology Information URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313618/> (дата обращения: 30.11.2022).
9. Лекция 1. Первичный анализ речевых сигналов. // Alpha Cephei speech Recognition URL: <https://alphacepei.com/ru/research> (дата обращения: 12.12.2022).
10. Панфилов И.А., Алексеев М.С., Сивцова Е.И. Извлечение признаков голосового корсега // Цифровая трансформация экономических систем:

- проблемы и перспективы (экопром-2022). - СПб: ПОЛИТЕХ-ПРЕСС, 2022. - С. 794-796.
11. *El Ayadi M., Kamel M. S., Karray F.* Survey on speech emotion recognition: Features, classification schemes, and databases //Pattern recognition. – 2011. – Т. 44. – №. 3. – С. 572-587.
 12. *Abbaschian B. J., Sierra-Sosa D., Elmaghraby A.* Deep learning techniques for speech emotion recognition, from databases to models //Sensors. – 2021. – Т. 21. – №. 4. – С. 1249.
 13. *Swain M., Routray A., Kabisatpathy P.* Databases, features and classifiers for speech emotion recognition: a review //International Journal of Speech Technology. – 2018. – Т. 21. – С. 93-120.
 14. *Chen L. et al.* Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction //Information Sciences. – 2020. – Т. 509. – С. 150-163.

Егорчев Антон Александрович, кандидат технических наук, директор института вычислительной математики и информационных технологий, КФУ, Казань, ул. Кремлевская, 35. Email: anton@egorchev.ru.

Чикрин Дмитрий Евгеньевич, доктор технических наук, директор института искусственного интеллекта, робототехники и системной инженерии, КФУ, Казань, ул. Сайлиха Сайдашева д.12 к.3, Email: dmitry.kfu@ya.ru.

Пашин Дмитрий Михайлович, доктор технических наук, проректор по цифровой трансформации и инновационной деятельности, КФУ, Казань, ул. Кремлевская д.18. Email: dmitry.m.pashin@gmail.com.

Сарамбаев Никита Андреевич, без степени, аспирант института вычислительной математики и информационных технологий, КФУ, Казань, ул. Кремлевская д. 35. Email: sarambaev@gmail.com Тел: +79179029352/

Фахрутдинов Адель Фердинандович, без степени, аспирант института физики, КФУ, Казань, ул. Кремлевская д.16а Email: timvaz@yandex.ru. Тел: +79872394153/

Egorchev Anton Alexandrovich, Ph.D of technical sciences, Director of Institute Computational Mathematics and IT, KFU, Kazan, Kremlevskaya st. 35. Email: anton@egorchev.ru

Chikrin Dmitry Evgen'evich, Doctor of technical sciences, Director of Institute of Artificial Intelligence, Robotics and System Engineering, KFU, Kazan, Salikha Saidasheva st. 12, 3 enc. Email: dmitry.m.pashin@gmail.com.

Pashin Dmitry Mikhailovich, Doctor of technical sciences, Vice-Rector for Digital Transformation and Innovation, KFU, Kazan, Kremlevskaya st. 18. Email: dmitry.kfu@ya.ru.

Sarambaev Nikita Andreevich, no degree, graduate student, Institute of Computational Mathematics and IT, IT, KFU, Kazan, Kremlevskaya st. 35. Email: sarambaev@gmail.com. Tel: +79179029352

Fakhrutdinov Adel Ferdinandovich, no degree, graduate student, Institute of Physics, KFU, Kazan, Kremlevskaya st. 16a. Email: timvaz@yandex.ru. Tel: +79872394153