УДК 623.76
DOI 10.52575/2687-0932-2023-50-4-944-954

Автоматическое обнаружение гнева и агрессии в речевых сигналах

[1] Балабанова Т.Н., [2] Абрамов К.В., [3] Болдышев А.В., [4] Долбин Д.М.
[1] Белгородский государственный национальный исследовательский университет,
Россия, 308015, г. Белгород, ул. Победы, д. 85
[2] Московский технический университет связи и информатики,
Россия, 111024, г. Москва, ул. Авиамоторная, д. 8
[3] Белгородский филиал ПАО Ростелеком,
Россия, 308009, г. Белгород, просп. Б-Хмельницкого, д. 81
[4]Белгородский университет кооперации, экономики и права
Россия, 308023, г. Белгород, ул. Садовая, д. 116а
E-mail: sozonova@bsu.edu.ru, kirya_abramov_2002@bk.ru,
Aleksei_Boldyshev@center.rt.ru, DolbinDM@mail.ru

Аннотация. В статье рассматривается вопрос обнаружения гнева и агрессии в речевом сигнале. Рассмотрены принципиальные отличия гнева от агрессии. Проведен обзор решений распознавания деструктивного поведения в виде гнева и агрессии по речевому сигналу, представленных в различных современных публикациях. Рассмотрены основные методы классификации, используемые для решения задачи распознавания эмоций по речи. Проанализировано информационное обеспечение в виде русскоязычных и нерусскоязычных речевых баз данных, применяемых для тренировки моделей при распознавании эмоций. Сформулированы основные проблемы использования речевых баз данных. Рассмотрен вопрос выбора параметров речевого сигнала, используемых для классификации эмоций в общем и деструктивном поведении в частности. Реализовано распознавание гнева на русскоязычной базе данных Dusha с использованием двух подходов тремя методами классификации.

Ключевые слова: речевые данные, речевые базы данных, классификация, методы классификации, низкоуровневые дескрипторы, распознавание гнева, распознавание агрессии

Для цитирования: Балабанова Т.Н., Абрамов К.В., Болдышев А.В., Долбин Д.М. 2023. Автоматическое обнаружение гнева и агрессии в речевых сигналах. Экономика. Информатика, 50(4): 944-954. DOI: 10.52575/2687-0932-2023-50-4-944-954

Automatic Detection of Anger and Aggression in Speech Signals

[1] Tatiana N. Balabanova, [2] Kirill V. Abramov, Alexey V. Boldyshev, Dmitry M. Dolbin
[1] Belgorod State National Research University,
85 Pobedy Str., Belgorod, 308015, Russia
[2] Moscow Technical University of Communication and Informatics,
8 Aviamotornaya Str, Moscow, 111024, Russia
[3] Belgorod branch of PJSC Rostelecom,
81 B-Khmelnitsky Av., Belgorod, 308009, Russia
[4] Belgorod University of Cooperation, Economics and Law,
116a Sadovaya str., Belgorod, 308023, Russia
E-mail: sozonova@bsu.edu.ru, kirya_abramov_2002@bk.ru,
Aleksei_Boldyshev@center.rt.ru, DolbinDM@mail.ru

Abstract. The article discusses the issue of detecting anger and aggression in a speech signal. The fundamental differences between anger and aggression are considered. A review of solutions for

recognizing destructive behavior in the form of anger and aggression using a speech signal, presented in various modern publications, was carried out. The main classification methods used to solve the problem of recognizing emotions from speech are considered. Information support in the form of Russian-language and non-Russian-language speech databases used to train models for recognizing emotions is analyzed. The main problems of using speech databases are formulated. The issue of choosing speech signal parameters used to classify emotions in general and destructive behavior in particular is considered. Implemented anger recognition on the Russian-language Dusha database using two approaches and three classification methods.

Keywords: speech data, speech databases, classification, classification methods, low-level descriptors, anger recognition, aggression recognition

For citation: Balabanova T.N., Abramov K.V., Boldyshev A.V., Dolbin D.M. 2023. Automatic Detection of Anger and Aggression in Speech Signals. Economics. Information technologies, 50 (4): 944-954 (in Russian). DOI: 10.52575/2687-0932-2023-50-4-944-954

Введение

В настоящее время решение задачи распознавания эмоций человека по речевому сигналу является довольно востребованной в различных областях жизнедеятельности человека. Распознавание эмоций и других невербальных проявлений в речевом сигнале осуществляется посредством паралингвистического анализа. То есть, паралингвистика рассматривает речь с точки зрения того, как она произносится, а не того, что конкретно произносится. Так, системы паралингвистического анализа речи используются для определения удовлетворенности клиентов в колл-центрах с целью определения вероятности ложных сообщений в банковских системах и приеме на работу. Относительно новой областью является использование паралингвистического анализа речи для обеспечения безопасности. С этой точки зрения обнаружение деструктивного поведения человека по речи может быть востребовано в различных сферах, например, обнаружение гнева или агрессии в колл-центрах (как со стороны оператора, так и со стороны клиента); на предприятиях (особенно стратегического направления); в торговых центрах, автовокзалах, аэропортах и даже просто на улице. Следует отметить, что использование аудиоаналитики является менее дорогостоящим по отношению к применению видеоаналитики. Стоимость микрофона меньше, чем камер видеонаблюдения. К тому же дополнительное использование аудиоаналитики позволит более эффективно осуществлять распознавание диструктивного поведения, выражающегося в виде гнева или агрессии.

Под агрессией в психологии понимается деструктивное поведение, которое может привести к физическому насилию и причинению вреда как себе, так и окружающим. При агрессивном поведении человек является мотивированным и, как правило, считает, что он поступает обоснованно и верно. В работе [Кажберова, Чхартишвили, Губанов, Козицин, Белявский, Федянин, Черкасов, Мешков, 2023] дано следующее определение агрессии: агрессия – специфическая форма речевого поведения (в том числе в письменной речи), которая мотивирована аффективным состоянием говорящего. По классификации Басса существуют различные формы агрессивного поведения [Buss, 1957]. В первом приближении Басс выделяет активную и пассивную агрессию, каждая из которых, в свою очередь, делится на физическую и вербальную. Нанесение оскорблений, злословие относится по классификации Басса к вербальной активной агрессии, которая зачастую предшествует физической активной агрессии. Поэтому обнаружение агрессии в речи человека представляется важным с точки зрения предотвращения физического проявления агрессии.

Следует отметить, что очень часто рассматривают как синоним понятия «гнев» и «агрессия». Однако гнев представляет собой эмоцию, которая возникает при сильном недовольстве и может выражаться в виде раздражения, злости, ярости. Агрессия – представляет собой результат гнева, выраженный в активном действии. То есть, гнев – это тип чувства, тогда как агрессия – это тип поведения.

Современное состояние вопроса

Поэтому при разработке систем обнаружения агрессии по речевому сигналу представляется целесообразным также рассматривать задачу распознавания гневных высказываний как возможных предшественников проявления агрессии.

Разработка паралингвистических систем обнаружения диструктивного поведения по речи ведется сравнительно недавно и довольно сложно оценить качество предложенных решений, поскольку разработчики, зачастую, не предоставляют полную информацию о своих решениях, а также методах и способах тестирования своих систем. Как правило, информация ограничивается новостными анонсами, рекламой и общими фразами о работе алгоритма. Однако для полноты картины рассмотрим некоторые решения, представленные в различных публикациях.

Технология SaluteSpeech от Сбербанка. Данная технология, помимо распознавания речи, анонсирует возможность распознавания эмоций по речевому сигналу. Разделение эмоции осуществляется на позитивную, нейтральную, негативную.

Компания Louroe, известная в области аудионаблюдения совместно с Sound Intelligence разработала программный продукт, который помимо взрывов, разбитого стекла и автомобильной сигнализации способен распознать агрессию в голосе человека.

Проект «AudioAnalytics», который разрабатывается в Великобритании, позволяет получить аналитику аудиосигнала и определить по ней тревожные события в виде сигнализации автомобиля, разбивающегося стекла, выстрела, крика.

Российская система аудиоаналитики «SistemaSarov» позволяет по аудиопотоку выделять артефакты и осуществлять их предварительную классификацию по уровню тревожности.

Таким образом, в настоящее время существует потребность в разработке и исследовании методов, алгоритмов и создании систем распознавания диструктивного поведения в виде гнева и агрессии по речевому сигналу.

Задача распознавания гнева и агрессии относится к задаче классификации по определенным признакам. В настоящее время задача классификации может решаться двумя группами методов: классические методы и нейросетевые (рисунок 1).

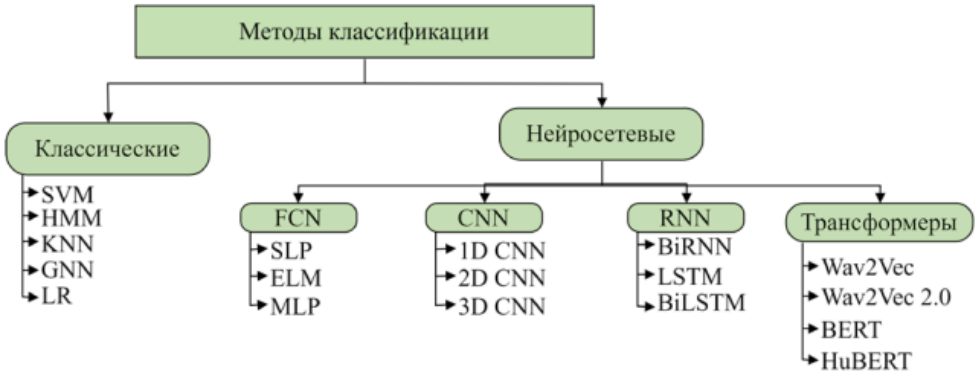


Рис. 1. Методы классификации
Fig. 1. Classification methods

Для распознавания эмоций человека (в том числе агрессию и гнев) по речевому сигналу чаще всего используются такие классические методы, как:

– метод k-ближайших соседей (K-Nearest Neighbor, KNN) [Dellaert, Polzin, Waibel, 1996];

– обобщённый метод моментов (Generalized Method of Moments, GMM) [Neiberg, Elenius, Laskowski, 2006];

– метод опорных векторов (Support Vector Machine, SVM) [Schuller, Batliner, Bergler, 2021];

– случайный лес (Random forest, RF);

– стохастический градиентный спуск (Stochastic Gradient Descent, SGD);

– скрытые марковские модели (англ. Hidden Markov Model, HMM) [Nogueiras, Moreno, Bonafonte, 2001];

– линейный дискриминантный анализ (Linear Discriminant Analysis, LDA) (Рисунок 1).

Данные методы неплохо зарекомендовали себя в решении задачи классификации и до сих пор используются для классификации эмоций человека по речевым данным. В качестве недостатка классических методов можно указать достаточно долгое обучение при использовании большого объема данных. Однако для распознавания агрессии классические методы классификации могут быть эффективно использованы ввиду небольшого количества речевого материала агрессии в имеющихся в настоящее время речевых базах данных.

Нейросетевые методы классификации при решении задач паралингвистики и распознавания диструктивного поведения по речевому сигналу зарекомендовали себя с положительной стороны. Первоначально использовались полносвязные нейронные сети (англ. Fully Connected Network, FCN) в виде однослойного (англ. Single-Layer Perceptron, SLP) [Raudys, 2003] и многослойного (англ. MultiLayer Perceptron, MLP) [Kruse, Borgelt, Klawonn, 2022] перцептрона. При появлении глубоких нейронных сетей (англ. Deep Neural Networks, DNN) [Sainath, Vinyals, Senior, 2015] многие паралингвистические разработки стали базироваться на них. В настоящее время широкое распространение в области паралингвистики получили архитектуры, основанные на сверточных (англ. Convolutional Neural Networks, CNN) и рекурентных (англ. Recurrent Neural Networks, RNN) нейронных сетях [Kim, Truong, Englebienne, 2017].

Информационное обеспечение

Одной из задач, при разработке систем распознавания диструктивного поведения в виде агрессии и гнева по речевому сигналу, является выбор информационного обеспечения в виде баз речевых данных. В настоящее время существует большое количество баз данных для распознавания эмоций по речевому сигналу на разных языках. Однако хочется заметить, что русскоязычных речевых баз данных для паралингвистического анализа не так много. В частности, имеется три базы данных: RAMAS, RUSLANA, Dusha.

База речевых данных RAMAS [Perepelkina, Kazimirova, Konstantinova, 2018]. Названная база речевых данных была получена в лабораторных условиях путем записи (как аудио модальности, так и видео) игры актеров. Всего в формировании базы участвовали пять пар актеров (50% – женщины, 50% – мужчины). Разметка эмоций по времени осуществлялась 21 экспертом с учетом мнения актеров о своем эмоциональном состоянии. База данных содержит 581 запись общей длительностью 7 часов. Разделение осуществляется по 7 эмоциям: гнев, печаль, отвращение, счастье, страх, удивление, нейтральное состояние.

База данных аффективной речи на русском языке RUSLANA. Эта база данных, как и RAMAS, была создана в лабораторных условиях. В ее создании принимали участие 61 человек (12 мужчин, 49 женщин). Каждый из дикторов прочитал 10 предложений, изображающих следующие эмоции: удивление, счастье, гнев, печаль, страх, нейтральное состояние. Всего в базе 3660 записей.

Самый большой открытый датасет для распознавания эмоций в устной речи на русском языке Dusha. Данная база данных состоит из 300000 аудиозаписей. Общая длительность 350 часов. База разделена на две части: Crowd – часть аудиоданных, полученная лабораторно путем имитации эмоции дикторами; Podcast – часть, полученная из реальных разговоров. Разделение эмоций осуществляется на 4 класса: позитив, грусть, злость/раздражение (гнев), нейтраль.

Несложно заметить, что с использованием русскоязычных баз речевых данных можно проводить обучение системы и в принципе эксперименты только по распознаванию таких деструктивных эмоций, как злость и гнев. Распознавание агрессии, как крайней формы проявления гнева или злости, по русскоязычным базам данных не представляется возможным.

Базы данных, использующиеся для обучения и тестирования систем распознавания агрессии, являются многомодальными и содержат помимо речевой информации визуальную. Основные характеристики баз данных для распознавания агрессии приведены в таблице 1.

Таблица 1
Table 1

Базы данных агрессивного поведения
Databases of aggressive behavior

Название	Объем данных, часов	Кол-во записей	Уровни	Кол-во дикторов	Язык
TR	0,6	н/и	3 уровня агрессии	н/и	Нидерландский
SD	0,5	8	3 уровня агрессии 5 уровней стресса	9	Английский нидерландский
NAA	Н/и	2240	5 уровней агрессии, страха, интенсивности 9 уровней валентности	16	Нидерландский

Материалы баз данных TR [Lefter, Rothkrantz, Burghouts, 2011] и SD [Lefter, Burghouts, Rothkrantz, 2014] являются натурными, а базы NAA [Lefter, Jomker, Tuente, 2017] лабораторными. Но во всех трех базах эмоции являются спонтанными, не наигранными. Однако при использовании этих баз данных основными проблемами являются следующие:

– нельзя быть точно уверенным, что представленный фрагмент речи относится к агрессивному поведению, так как часто даже эксперты-разметчики баз путают агрессию с гневом или злостью;

– при обучении и тестировании системы на нидерландском или английском языке нельзя быть уверенным, что качество ее работы не упадет при распознавании агрессии по русскоязычной речи, так как в работах [Makarova, 2000] указано, что выражение эмоций на русском языке имеет как универсальные, так и специфические черты выражения эмоций и аффектов.

Выбор параметров и распознавание гнева

Еще одной важной задачей при паралингвистическом анализе является выбор параметров. Для паралингвистического анализа используются различные акустические признаки речевого сигнала, которые можно разделить на два класса: экспертные и нейросетевые. В данной работе используются экспертные акустические признаки. Для общего описания фраз, а не отдельных фонем или слогов в паралингвистике используются низкоуровневые дискрипторы LLD (Low Level Descriptors). Условно дискрипторы LLD можно разделить на: энергетические, просодические, вокализованные и спектральные. Одним из инструментов, позволяющих вычислять данные дискрипторы для произвольного количества речевых данных, является бесплатный программный продукт openSmile [Eyben, Weninger, Gross, 2013]. Де-факто он является стандартом в компьютерной паралингвистике и,

в зависимости от выбора пользователя, позволяет получить 65 базовых LLD признака и большое количество суперсегментных признаков. Набор признаков зависит от выбранной конфигурации в openSmile.

В данной работе используется конфигурация INTERSPEECH 2009. Которая позволяет вычислить 16 низкоуровневых дескриптора (LLD) и 16 соответствующих коэффициентов дельта-регрессии. К этим 32 дескрипторам применено 12 функционалов, что дает всего 384 признака.

Основные дескрипторы конфигурации INTERSPEECH 2009:

– мел-кепстральные коэффициенты (Mel-Frequency Cepstral Coefficient, MFCC),

– среднеквадратическая энергия (Root Mean Square, RMS),

– скорость перехода через ноль (Zero-Crossing Rate, ZCR),

– частота основного тона (F_0),

– оценка автокорреляции (Autocorrelation Based Estimation).

Данная конфигурация была использована, поскольку содержит относительно небольшое количество LLD, что является важным для реализации классификации эмоций классическими методами.

В качестве методов для распознавания деструктивного поведения человека (в частности, гнева) по речевому сигналу были использованы три классических метода классификации:

1. Методах k-ближайших соседей (KNN),
2. Случайный лес (RF),
3. Стохастический градиентный спуск (SGD).

В работе осуществлялось распознавание гнева, поскольку целью было использование русскоязычных баз данных. В качестве экспериментальной речевой базы была выбрана Dusha, как самая большая русскоязычная эмоциональная речевая база.

В процессе организации экспериментов были использованы два подхода:

Подход 1 заключался в классификации всех имеющихся четырех эмоций: позитив, грусть, злость/раздражение (гнев), нейтраль.

Подход 2 заключался в классификации речевого сигнала на два класса. Первый класс – гнев, второй класс – все остальные эмоции.

Для обучения и тестирования моделей были использованы два набора данных: Train – обучающая выборка, test – тестовая выборка. Для соблюдения баланса между объектами разных классов размеры тренировочной и тестовой выборок для второго подхода были уменьшены. Объемы выборок для обучения и тестирования по всем трем моделям представлены в таблице 2.

Таблица 2
Table 2

Объем данных для метода KNN
Data volume for the KNN method

№	Модель	Кол-во записей тренировочной выборки, шт.	Кол-во записей тестовой выборки, шт.	Кол-во записей «гнев» в test, шт.
1	Подход 1	11696	2924	731
2	Подход 2	5848	1462	731

Оценка качества работы методов, используемых для распознавания гнева по речевому сигналу, осуществлялась по показателю невзвешенной средней полноты (Unweighted Average Recall, UAR). Этот показатель является наиболее распространенным для оценки качества распознавания [Величко, 2022].

$$UAR = \frac{1}{k}\sum_{i=1}^{k}\frac{N_c^{(i)}}{N_0^{(i)}}, \quad (1)$$

где $N_c^{(i)}$ – количество верно распознанных элементов i-го класса, $N_0^{(i)}$ – общее количество элементов i-го класса, N – общее количество объектов, k – количество классов.

1. Метод k-ближайших соседей (KNN).

В данном методе оба подхода были реализованы для параметра k=1,…, 100.

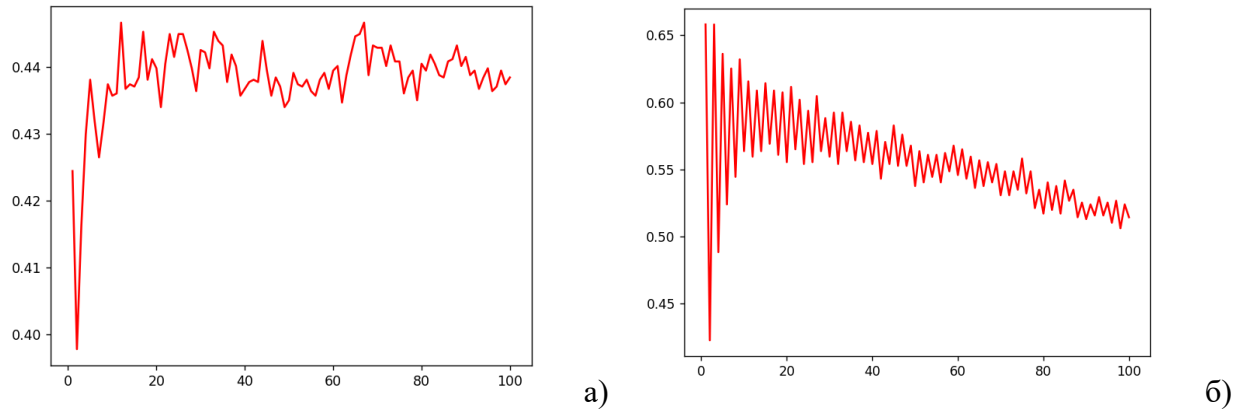
Результаты эксперимента представлены в таблице 3.

Таблица 3
Table 3

UAR метода KNN при оптимальных результатах
UAR of the KNN method with optimal results

№	Модель	k_{optim}	UAR
1	Подход 1	12	0,45
2	Подход 2	3	0,69

На рисунке 2 представлены графики изменения показателя UAR для каждого подхода при k=1,…,100.



а) б)

Рис. 2. UAR метода KNN при k=1,…,100 а) Подход 1, б) Подход 2
Fig. 2. UAR of the KNN method for k=1,…,100 a) Approach 1, b) Approach 2

По результатам эксперимента можно заключить, что из двух представленных моделей наилучший результат показала модель 2 при параметре k=3. Однако максимальное значение показателя UAR равно 0,69, то есть из 100 речевых сигналов, содержащих эмоцию «гнев», распознано было 69.

2. Случайный лес (RF).

В данном методе были использованы также два подхода.

Параметры, которые были использованы для реализации метода и их значения:

n_estimators – число деревьев в «лесу»: [10, 50, 100, 200],

criterion – критерий для разбиения выборки в вершине: ['gini', 'entropy', 'log_loss'],

max_depth – максимальная глубина дерева: [50, 100, 200, None],

min_samples_split – минимальное количество выборок, необходимых для разделения внутреннего узла: [2, 10, 20, 50, 100],

min_samples_leaf – минимальное количество выборок, которое требуется для конечного узла: [1, 5, 10, 15],

max_features – число признаков, по которым ищется разбиение: ['sqrt', 'log2'].

Таким образом было обучено 1920 вариантов моделей, которые полностью перебирают все вышеупомянутые параметры.

Оценка качества распознавания методом случайный лес представлена в таблице 4.

Таблица 4
Table 4

UAR метода случайный лес при оптимальных результатах
UAR of the random forest method with optimal results

№	Модель	Значение параметров оптимального результата	UAR
1	Подход 1	n_estimators = 200 criterion = log_loss max_depth = 100 min_samples_split = 2 min_samples_leaf = 10 max_features = sqrt	0,47
2	Подход 2	n_estimators = 100 criterion = entropy max_depth = 100 min_samples_split = 20 min_samples_leaf = 5 max_features = sqrt	0,73

По результатам эксперимента можно заключить, что метод случайный лес дал лучший результат для подхода 2 то есть при классификации на 2 класса. В сравнении с методом KNN, случайный лес показал немного лучший результат как для подхода 1, так и для подхода 2.

3. Стохастический градиентный спуск (SGD).

В данном эксперименте были использованы также два подхода. Максимальное количество эпох при обучении составило 20000.

Параметры, которые были использованы для реализации метода:

loss – используемая функция потерь: ['hinge', 'log_loss', 'modified_huber', 'squared_hinge', 'perceptron', 'squared_error', 'huber', 'epsilon_insensitive', 'squared_epsilon_insensitive'],

penalty – метод регуляризации: ['l2', 'l1', 'elasticnet', None]?

alpha – параметр регуляризации: [0.1, 0.01, 0.001, 0.0001, 0.00001, 0.000001].

Таким образом было обучено 216 моделей, которые полностью перебирают все вышеупомянутые параметры.

Оценка качества распознавания методом стохастический градиентный спуск представлена в таблице 5.

Таблица 5
Table 5

UAR метода стохастический градиентный спуск при оптимальных результатах
UAR of the stochastic gradient descent method with optimal results

№	Модель	Значение параметров оптимального результата	UAR
1	Подход 1	loss = log_loss penalty = elasticnet alpha = 0,01	0,49
2	Подход 2	loss = log_loss penalty = elasticnet alpha = 0,01	0,73

По данным, приведенным в таблице, видно, что метод стохастического градиентного спуска дал такой же результат, как и метод случайного леса. Однако с вычислительной точки зрения метод стохастического градиентного спуска гораздо быстрее, что может быть аргументом использования его, а не метода случайного леса.

Все результаты эксперимента не противоречат результатам распознавания агрессии на англоязычных базах речевых данных [100].

С целью повышения качества распознавания гнева по речевому сигналу были проведены эксперименты с использованием всех трех методов распознавания. Решающее правило по отнесению фрагмента речевого сигнала к классу «гнев» осуществлялось путем сравнения двух показателей: DR_1 и DR_2.

$$DR_1 = \sum_{i=1,\ if\ m_i=1}^{3} UAR_i, \quad (2)$$

$$DR_2 = \sum_{i=1,\ if\ m_i=0}^{3} (1 - UAR_i), \quad (3)$$

где m_i – результат распознавания гнева i-м методом i=1,2,3, UAR_i – показатель невзвешенной средней полноты i-го метода.

Если $DR_1 > DR_2$, то речевой сигнал относится к классу «гнев», в противном случае – к классу «не гнев».

Заключение и направление дальнейших исследований

Эксперименты по распознаванию гнева, проведенные с использованием трех методов классификации, позволили повысить оценку *UAR* распознавания до 0,76. Это говорит о том, что используемые методы делают ошибку приблизительно на одних и тех же файлах. Метод KNN дает худший результат из рассмотренных методов классификации.

Исходя из этого, в качестве дальнейших исследований предполагается анализ влияния выбора параметров различных конфигураций openSmile на качество распознавания гнева и агрессии. А также выявление параметров агрессивного речевого сигнала, не зависящих от языка говорящего.

В качестве способов распознавания предполагается применение нейросетевых методов оценки эмоций для распознавания гнева и агрессии.

Список литературы

Величко А.Н. 2022. Метод анализа речевого сигнала для автоматического определения агрессии в разговорной речи. Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. № 4. С. 180-188.

Кажберова В.В., Чхартишвили А.Г., Губанов Д.А., Козицин И.В., Белявский Е.В., Федянин Д.Н., Черкасов С.Н., Мешков Д.О. 2023. Агрессия в общении медиапользователей: анализ особенностей поведения и взаимного влияния Вестник Московского университета. Серия 10: Журналистика. № 3. С. 26-56.

Buss A., Durkee A.An inventory for assessing different kinds of hostility. 1957. Journal of Consulting Psychology. 21(4): 343–349. URL: https://doi.org/10.1037/h0046900.

Dellaert F., Polzin T., Waibel A. 1996. Recognizing emotion in speech. Proceedings of the 4th Int. Conf. Spoken Lang. Process (ICSLP). pp. 1970–1973.

Eyben F., Weninger F., Gross F., et al. 2013. Recent developments in opensmile, the munich open-source multimedia feature extractor. Proceedings of ACM International Conference on Multimedia. pp. 835–838.

Kim J., Truong K.P., Englebienne G., et al. 2017. Learning spectro-temporal features with 3D CNNs for speech emotion recognition. Proceedings of the 7th International Conference on Affective Computing and Intelligent Interaction (ACII). pp. 383–388.

Kruse R., Borgelt C., Klawonn F., et al. 2022. Multi-layer perceptrons. Computational Intelligence. Springer, Cham. pp. 53-124.

Lefter I., Burghouts G.J., Rothkrantz L.J.M. 2014. An audio-visual dataset of human–human interactions in stressful situations. Journal on Multimodal User Interfaces. 8(1): 29-41.

Lefter I., Jomker C.M., Tuente S.K., et al. 2017. NAA: A multimodal database of negative affect and aggression. Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE. pp. 21-27.

Lefter I., Rothkrantz L.J.M., Burghouts G., et al. 2011. Addressing multimodality in overt aggression detection. Proceedings of the International Conference on Text, Speech and Dialogue. Springer, Berlin, Heidelberg. pp. 25-32.

Makarova V. 2000. Acoustic cues of surprise in Russian questions. Journal of the Acoustical Society of Japan (E), 21 (5): 243-250.

Neiberg D., Elenius K., Laskowski K. 2006. Emotion recognition in spontaneous speech using GMMs. Proceedings of the 9th Int. Conf. Spoken Lang. Process. pp. 809– 812.

Nogueiras A., Moreno A., Bonafonte A., et al. 2001. Speech emotion recognition using hidden Markov models. Proceedings of the 7th Eur. Conf. Speech Commun. Technol. pp. 746–749.

Perepelkina O., Kazimirova E., Konstantinova M. 2018. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing. Proceedings of the International Conference on Speech and Computer. Springer, Cham. pp. 501-510.

Raudys Š. 2003. On the universality of the single-layer perceptron model. Neural Networks and Soft Computing. Physica. Heidelberg. pp. 79-86.

Sainath T.N., Vinyals O., Senior A., et al. 2015. Convolutional, long short-term memory, fully connected deep neural networks. Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4580–4584.

Schuller B.W., Batliner A., Bergler C., et al. 2021. The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primates. Proceedings of Interspeech. pp. 431–435.

References

Kazhberova V.V., Chkhartishvili A.G., Gubanov D.A., Kozitsin I.V., Belyavsky E.V., Fedyanin D.N., Cherkasov S. N., Meshkov D.O. 2023. Aggression in communication between media users: analysis of behavioral characteristics and mutual influence. Bulletin of Moscow University. Episode 10: Journalism. No. 3. pp. 26-56.

Velichko A.N. 2022. Method of speech signal analysis for automatic detection of aggression in spoken speech. Bulletin of Voronezh State University. Series: System analysis and information technologies. No. 4. pp. 180-188.

Buss A., Durkee A.An inventory for assessing different kinds of hostility. 1957. Journal of Consulting Psychology. 21(4): 343–349. URL: https://doi.org/10.1037/h0046900.

Dellaert F., Polzin T., Waibel A. 1996. Recognizing emotion in speech. Proceedings of the 4th Int. Conf. Spoken Lang. Process (ICSLP). pp. 1970–1973.

Eyben F., Weninger F., Gross F., et al. 2013. Recent developments in opensmile, the munich open-source multimedia feature extractor. Proceedings of ACM International Conference on Multimedia. pp. 835–838.

Kim J., Truong K.P., Englebienne G., et al. 2017. Learning spectro-temporal features with 3D CNNs for speech emotion recognition. Proceedings of the 7th International Conference on Affective Computing and Intelligent Interaction (ACII). pp. 383–388.

Kruse R., Borgelt C., Klawonn F., et al. 2022. Multi-layer perceptrons. Computational Intelligence. Springer, Cham. pp. 53-124.

Lefter I., Burghouts G.J., Rothkrantz L.J.M. 2014. An audio-visual dataset of human–human interactions in stressful situations. Journal on Multimodal User Interfaces. 8(1): 29-41.

Lefter I., Jomker C.M., Tuente S.K., et al. 2017. NAA: A multimodal database of negative affect and aggression. Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE. pp. 21-27.

Lefter I., Rothkrantz L.J.M., Burghouts G., et al. 2011. Addressing multimodality in overt aggression detection. Proceedings of the International Conference on Text, Speech and Dialogue. Springer, Berlin, Heidelberg. pp. 25-32.

Makarova V. 2000. Acoustic cues of surprise in Russian questions. Journal of the Acoustical Society of Japan (E), 21 (5): 243-250.

Neiberg D., Elenius K., Laskowski K. 2006. Emotion recognition in spontaneous speech using GMMs. Proceedings of the 9th Int. Conf. Spoken Lang. Process. pp. 809– 812.

Nogueiras A., Moreno A., Bonafonte A., et al. 2001. Speech emotion recognition using hidden Markov models. Proceedings of the 7th Eur. Conf. Speech Commun. Technol. pp. 746–749.

Perepelkina O., Kazimirova E., Konstantinova M. 2018. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing. Proceedings of the International Conference on Speech and Computer. Springer, Cham. pp. 501-510.

Raudys Š. 2003. On the universality of the single-layer perceptron model. Neural Networks and Soft Computing. Physica. Heidelberg. pp. 79-86.

Sainath T.N., Vinyals O., Senior A., et al. 2015. Convolutional, long short-term memory, fully connected deep neural networks. Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4580–4584.

Schuller B.W., Batliner A., Bergler C., et al. 2021. The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primates. Proceedings of Interspeech. pp. 431–435.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.
Conflict of interest: no potential conflict of interest related to this article was reported.

Поступила в редакцию 07.11.2023
Поступила после рецензирования 27.11.2023
Принята к публикации 01.12.2023

Received November 07, 2023
Revised November 27, 2023
Accepted December 01, 2023

ИНФОРМАЦИЯ ОБ АВТОРАХ

Балабанова Татьяна Николаевна, кандидат технических наук, доцент кафедры информационно-телекоммуникационных систем и технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

Абрамов Кирилл Владиславович, студент 4 курса факультета информационных технологий, Московский технический университет связи и информатики, г. Москва, Россия

Болдышев Алексей Владимирович, кандидат технических наук, ведущий инженер электросвязи, Белгородский филиал ПАО Ростелеком, г. Белгород, Россия

Долбин Дмитрий Михайлович, магистрант 2 курса факультета таможенного дела и информационных технологий, Белгородский университет кооперации, экономики и права, г. Белгород, Россия

INFORMATION ABOUT THE AUTHORS

Tatiana N. Balabanova, Candidate of Technical Sciences, Associate Professor of the Department of Information and Telecommunication Systems and Technologies, Belgorod State National Research University, Belgorod, Russia

Kirill V. Abramov, 4th year student of the Faculty of Information Technologies of the Moscow Technical University of Communications and Informatics, Moscow, Russia

Alexey V. Boldyshev, Candidate of Technical Sciences, Leading Telecommunications Engineer of the Belgorod branch of PJSC Rostelecom, Belgorod, Russia

Dmitry M. Dolbin, 2nd year master's student, Faculty of Customs Affairs and Information Technologies, Belgorod University of Cooperation, Economics and Law, Belgorod, Russia