



<http://dx.doi.org/10.35596/1729-7648-XXXX-XX-X-XX-XX>

Оригинальная статья

Original paper

УДК 004.934.2+534.784

РАСПОЗНАВАНИЕ РЕЧЕВЫХ ЭМОЦИЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДА ОПОРНЫХ ВЕКТОРОВ И НАДСЕГМЕНТНЫХ АКУСТИЧЕСКИХ ПРИЗНАКОВ

ВАШКЕВИЧ М.И., КРАСНОПРОШИН Д.В.

*Белорусский государственный университет информатики и радиоэлектроники
(г. Минск, Республика Беларусь)*

Поступила в редакцию

© Белорусский государственный университет информатики и радиоэлектроники, 2023

Аннотация. В данном исследовании изучается проблема распознавания речевых эмоций с использованием мел-частотных кепстральных коэффициентов (МЧКК) при помощи классификатора на основе метода опорных векторов (МОВ). Для проведения экспериментов использовался набор данных RAVDESS. Предложена модель, которая использует 306-компонентный вектор признаков, вычисленный на основе МЧКК, в качестве входных данных для классификатора МОВ. Для оценки качества модели использовался невзвешенное среднее значение полноты (UAR). Эксперименты проводились с различными функциями ядра для МОВ (например, линейный, полиномиальный и радиальный базис) и разным размером кадра для извлечения МЧКК (от 23 до 185 мс). Результаты экспериментов демонстрируют многообещающую точность (UAR = 48%), демонстрируя потенциал этого подхода для таких приложений, как голосовые помощники, виртуальные агенты и диагностика психического здоровья.

Ключевые слова: голосовой сигнал, МЧКК, извлечение аудио признаков, распознавание, машинное обучение.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Для цитирования. Вашкевич М.И., Краснопрошин Д.В. Распознавание речевых эмоций с использованием метода опорных векторов и надсегментных акустических признаков. Доклады БГУИР. 2023; **(*): ***-***.

SPEECH EMOTION RECOGNITION USING SVM CLASSIFIER WITH SUPRASEGMENTAL MFCC FEATURES

MAXIM.I. VASHKEVICH, DANIIL V. KRASNOPROSHIN

*Belarusian state university of informatics and radioelectronics
P.Brovki str., 6, Minsk, 220013, Republic of Belarus*

Submitted

© Belarusian State University of Informatics and Radioelectronics, 2023

Abstract. This study explores speech emotion recognition (SER) using mel-frequency cepstral coefficients (MFCCs) and Support Vector Machines (SVMs) classifier on the RAVDESS dataset. We proposed a model which uses 306-component suprasegmental MFCC feature vector as an input downstream by SVM classifier. To evaluate the quality of the model, unweighted average recall (UAR) was used. We evaluate different kernel function for SVM (such as linear, polynomial and radial basis) and different frame size for MFCC extraction (from 23 to 185 ms). Experimental results demonstrate promising accuracy (UAR = 48%), showcasing the potential of this approach for applications like voice assistants, virtual agents, and mental health diagnostics.

Keywords: voice signal, MFCC, DSP, audio feature extraction, recognition, machine learning.

Conflict of interests. The authors declare no conflict of interests.

For citation. Vashkevich M.I., Krasnoproshin D.V., Speech emotion recognition using SVM classifier with suprasegment MFCC features. Doklady BGUIR. 2023; **(*): ***-***.

Введение

Область распознавания эмоций по речи быстро развивается в последние десятилетия благодаря росту производительности вычислительных систем и широкому интересу к этому вопросу исследователей в области психологии, психиатрии и информатики [1], [2]. Эмоции часто влияют на процессы принятия решений, поэтому распознавание эмоций может представлять интерес для построения более эффективного общения, включая диалоговые системы (голосовые помощники, чат-боты). Задача распознавания негативных эмоций, таких как стресс, гнев, усталость является важным аспектом с точки зрения обеспечения безопасности дорожного движения при использовании интеллектуальных транспортных средств, поскольку позволяет им реагировать на эмоциональное состояние водителя [3].

В данной работе рассматривается задача определения эмоций на основе анализа речевого сигнала. Одна из основных проблем данного подхода связана с определением набора признаков, эффективно описывающих эмоциональное состояние диктора [1], [4-6].

Для построения системы распознавания эмоций авторами статьи предлагается использовать мел-частотные кепстральные коэффициенты (МЧКК) [7] для получения признаков и метод опорных векторов (МОВ) [7] в качестве метода классификации.

Признаки на основе МЧКК широко применяются в системах распознавания эмоций по речи [1, 5]. МЧКК воспроизводят реакцию слуховой системы человека на звук, улавливая соответствующую акустическую информацию [8]. Формируя представление аудиосигнала в частотной области, МЧКК уменьшают размерность данных, сохраняя при этом важные особенности, что делает его пригодным для алгоритмов машинного обучения, таких как МОВ.

В свою очередь МОВ является простым и надежным подходом к задаче классификации, который обеспечивает адаптируемость к многомерным пространствам признаков. МОВ основан на принципе поиска оптимальной гиперплоскости, максимально разделяющей разные классы в пространстве признаков [9]. В контексте распознавания эмоций по речи это означает, что МОВ может эффективно дифференцировать эмоциональные состояния [5]. Кроме того, МОВ может учитывать нелинейные отношения с помощью функций ядра (*kernel function*), что позволяет улавливать сложные закономерности в речевых данных.

Извлечение речевых признаков

Первым этапом системы по распознаванию эмоций по речи является предварительная обработка входных аудиоданных [1, 5]. В данной работе речевые признаки рассчитывались на основании МЧКК [8]. Расчет МЧКК относится к методам кратковременного анализа речевого сигнала, которые предполагают разбиение сигнала на фреймы (короткие сегменты). Считается,

что в интервале от 10 до 30 мс голосовой сигнал можно считать стационарным. На рис. 1 представлена схема вычисления МЧКК.

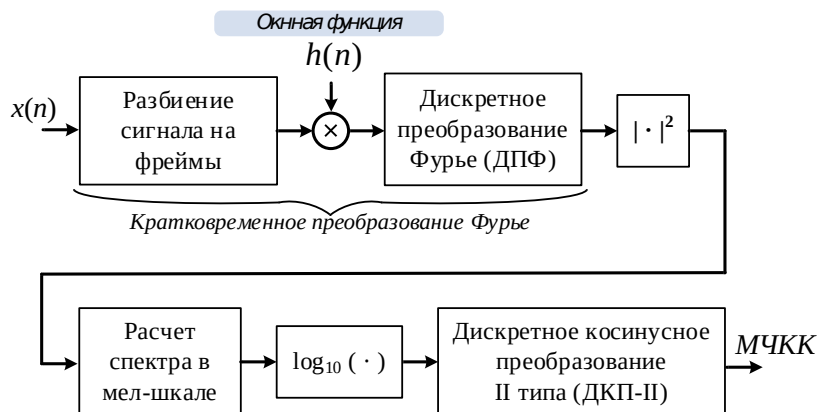


Рис. 1. Схема вычисления мел-частотных кепстральных коэффициентов (МЧКК)
Fig. 1. Scheme for calculating mel-frequency cepstral coefficients (MFCC)

Согласно рис. 1 процесс извлечения МЧКК включает следующие шаги:

- 1) вычисление кратковременного преобразования Фурье (КВПФ) и нахождение квадрата модуля КВПФ для получения спектрограммы сигнала;
- 2) вычисление мел-спектрограммы (энергия сигнала из шкалы герц переводится в мел-шкалу, отражающую свойства человеческого слуха);
- 3) взятие логарифма от энергии сигнала в мел-частотных полосах;
- 4) применение декоррелирующего преобразования, в качестве которого используется дискретное косинусное преобразование второго типа (ДКП-II).

В качестве иллюстрации на рис. 2 показан пример речевого сигнала, выражающего эмоцию гнева.

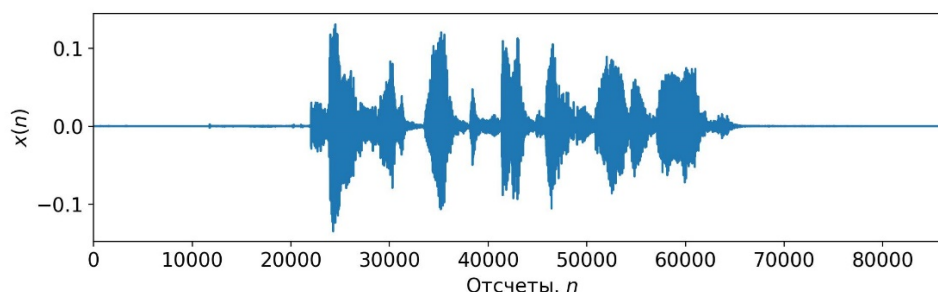
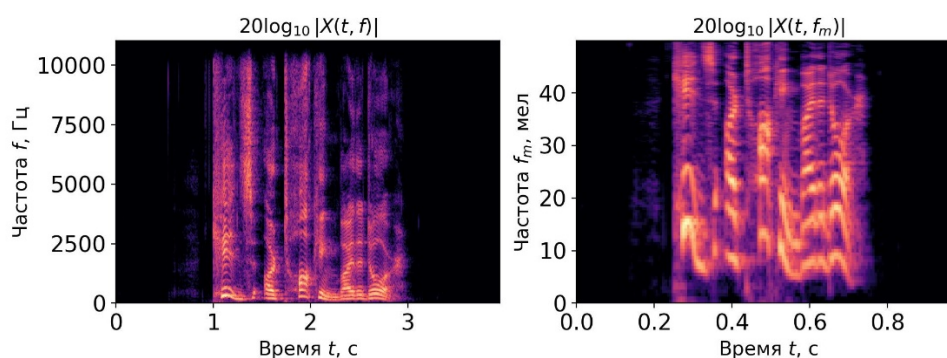


Рис. 2. Представление речевого сигнала, выражающего гнев («Kids are talking by the door»)
Fig. 2. Representation of the speech signal expressing anger

На рис. 3 представлен результат вычисления КВПФ и мел-спектрограммы сигнала, представленного на рис. 1.



а) б)

Рис. 3. речевого сигнала, выражающего гнев: а) спектрограмма (КВПФ), б) мелспектрограмма

Fig. 3. Speech signal expressing anger: a) spectrogram (STFT), b) mel-spectrogram

На рис. 4 показана временная последовательность MFCC, рассчитанная для сигнала, представленного на рис. 1.

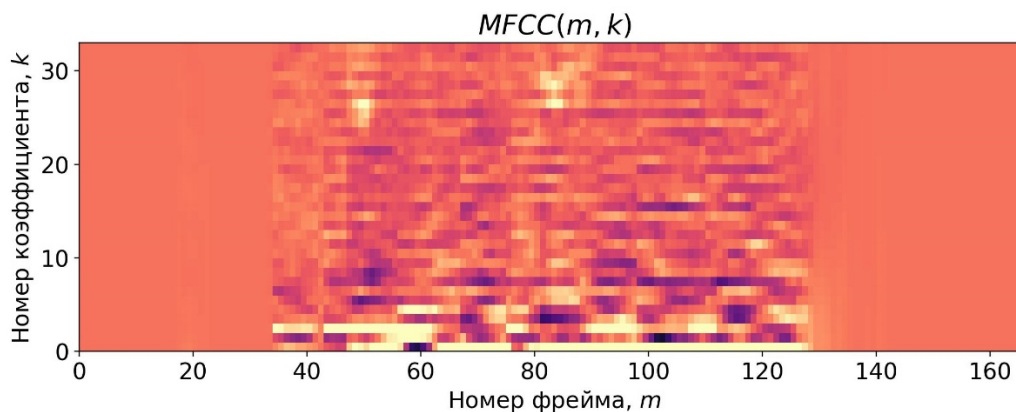


Рис. 4. Временная последовательность МЧКК

Fig. 4. Time-sequence of MFCC

В данной работе используются речевые сигналы с частотой дискретизации 48 кГц. Для обработки аудиосигналов использовалась библиотека librosa написанная на языке Python. КВПФ рассчитывается с использованием следующего набора размеров фреймов $N = \{1024, 2048, 4096, 8192\}$. Размер перекрытия между фреймами брался равным 50% от размера фрейма. Из каждого фрейма извлекалось $D=34$ МЧКК. После обработки одного аудиофайла мы получаем матрицу $MFCC(k, m)$ размером $D \times N_{frames}$, где N_{frames} – количество фреймов. Таким образом, m -й столбец матрицы представляет собой вектор МЧКК, вычисленный для временного фрейма с номером m .

На основании матрицы $MFCC$ рассчитывался вектор признаков фиксированной длины.

Чтобы получить единый вектор признаков для каждого аудиофайла, мы вычисляем средние $meanMFCC(k)$ (D признаков) и СКО – среднеквадратичные отклонения, $SDMFCC(k)$ – (D признаков) для МЧКК в матрице $MFCC$ вдоль оси времени. СКО может быть полезно для оценки степени изменчивости в МЧКК во времени.

По матрице МЧКК также производится извлечение первой и второй производных (первой и второй разности). Ниже поясняется смысл, который несет данная операция.

Первая производная МЧКК (Δ -МЧКК) представляет собой скорость изменения каждого коэффициента во времени. Таким образом, Δ -МЧКК отражает быстроту изменения спектральных характеристик звука во времени. Δ -МЧКК могут помочь в выявлении моментов, когда звук становится более высокочастотным или более громким, что может быть полезным для распознавания звуковых событий.

Вторая производная МЧКК (Δ^2 -МЧКК) представляет собой ускорение изменения каждого коэффициента во времени. Принятие во внимание Δ^2 -МЧКК позволяет более полно описать характер спектральных изменений речевого сигнала во времени.

Известно, что применение производных МЧКК позволяет улучшить производительность систем распознавания, классификации и детектирования речевых сигналов [1, 5]. Они могут использоваться для выделения важных характеристик аудиосигнала, таких как изменения в тональности, интонации, и ритме речи, что делает их полезными в приложениях, таких как распознавание речи, детектирование звуковых событий, и музыкальный анализ.

Для полученных матриц Δ -МЧКК и Δ^2 -МЧКК рассчитываются их средние значения и СКО, что в итоге дает $4 \times D = 136$ дополнительных признаков.

На основании матрицы МЧКК также были рассчитаны коэффициент асимметрии, эксцесс и межквантильный размах.

Коэффициент асимметрии (*skewness*) – это мера степени асимметрии распределения случайной величины. Она показывает, насколько сильно и в какую сторону смещено распределение относительно своего среднего значения. В работе для вычисления коэффициента асимметрии для каждого МЧКК:

$$SkMFCC(k) = \left(\frac{1}{N_{frames}} \right) \sum_{m=0}^{N_{frames}} \frac{(MFCC(k, m) - meanMFCC(k))^3}{SDMFCC(k)^3}. \quad (1)$$

В контексте анализа МЧКК коэффициент асимметрии может дать информацию о форме распределения и подчеркнуть определенные аспекты вариативности звукового сигнала.

Эксцесс (*kurtosis*) – это мера формы распределения случайной величины, которая показывает, насколько оно остроконечное или плоское по сравнению с нормальным распределением. В работе для вычисления эксцесса использовалась формула моментов:

$$kurtosisMFCC(k) = \left(\frac{1}{N_{frames}} \right) \sum_{m=0}^{N_{frames}} \frac{(MFCC(k, m) - meanMFCC(k))^4}{SDMFCC(k)^4 - 4}. \quad (2)$$

Межквартильный размах (*Interquartile Range, IQR*) – это мера разброса данных, которая используется для измерения разницы между верхним и нижним квартилями. Она показывает разброс значений в данных. При вычислении *IQR* выполняются шаги: 1) упорядочиваются данные по возрастанию; 2) находится значение первого квартиля (Q_1), которое отделяет нижние 25% наблюдений; 3) находится значение третьего квартиля (Q_3), которое отделяет нижние 75%; 4) вычислите *IQR*, как разницу между значениями Q_3 и Q_1 :

$$IQR = Q_3 - Q_1. \quad (3)$$

В контексте решаемой задачи *IQR* может быть полезен для измерения разброса в параметрах МЧКК.

Таким образом, в итоговый набор признаков были включены среднее значение МЧКК (34 признака), среднее квадратичное отклонение МЧКК (34 признака), среднее от первой и второй производных от МЧКК (2×34 признака), их среднее квадратическое отклонение (2×34), а также коэффициент асимметрии, эксцесс и межквантильный размах (по 34 признака для каждой метрики соответственно). Таким образом, для каждого аудиофайла мы получаем 306-компонентный вектор надментных признаков MFCC.

Речевая база

При проведении исследования в качестве исходного набора данных использовался Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [10]. RAVDESS содержит записи 24 актеров (12 мужчин, 12 женщин). Каждый актер произнес 104 различных сообщения (60 речевых высказываний и 44 песенных). В рамках данной работы использована только часть набора RAVDESS, а именно RAVDESS Emotional speech audio. Эта часть содержит 1440 файлов в формате wav (16 бит, 48 кГц): 60 записей на каждого из 24-х актеров. Речевые эмоции включают выражения нейтральности, спокойствия, счастья, грусти, гнева, страха, удивления и отвращения. Все эмоциональные состояния, кроме нейтрального, озвучивались на двух уровнях эмоциональной громкости (нормальная и повышенная). Актеры повторяли каждую вокализацию дважды.

Описание эксперимента

При проведении выполнения классификации в системе распознавания эмоционального состояния по речи применялся метод опорных векторов (MOB). Данный метод выполняет

классификацию путем построения N -мерных линейной (или нелинейной) разделяющей гиперплоскости, которые оптимально разделяют данные на классы. В МОВ предусмотрено использование функции ядра для преобразования исходного набора данных в многомерное пространство признаков. И уже в новом пространстве признаков выполняется построение оптимальной разделяющей гиперплоскости. В данной работе использовались известные ядерные функции: 1) линейное; 2) полиномиальное; 3) гауссово ядро с радиальной базисной функцией (РБФ) [9].

Построение классификатора на опорных векторах с использованием перечисленных выше ядер осуществлялось с помощью библиотеки *sklearn* языка Python.

Для тестирования классификатора использовался метод перекрестной проверки по k блокам (*k-fold cross-validation*) [9], который заключается в следующем.

1) Разделить набор разбивается на k блоков;

2) В цикле для $i = 1, 2, \dots, k$ выполняются следующие операции:

– блок i устанавливается в качестве тестового набора данных (*test data*)

– оставшиеся блоки принимаются как тренировочные данные (*train data*)

– выполняется обучение модели классификатора на тренировочных и оценивается ее эффективность на тестовых данных

– сохранение результатов классификации для данных из тестового набора;

– сброс параметров модели до исходного состояния для следующей итерации.

3) Расчет оценки эффективности модели на основе сохраненных результатов классификации данных.

В данной работе данные были разбиты на блоки следующим образом (в скобках указаны номера актеров):

– блок 0: (2, 5, 14, 15, 16)

– блок 1: (3, 6, 7, 13, 18)

– блок 2: (10, 11, 12, 19, 20)

– блок 3: (8, 17, 21, 23, 24)

– блок 4: (1, 4, 9, 22)

Такой порядок разбиения был предложен в [13]. Выбранная стратегия заключается в том, что каждый блок должен содержать одинаковое количество случайно выбранных образцов для каждого класса. При этом должно выполняться условие, что каждый актер представлен либо обучающей, либо валидационной выборке, но не в обоих [13].

Для оценки качества модели было вычислено среднее арифметическое (невзвешенное) полноты (*unweighted average recall, UAR*). *UAR* – это показатель, используемый для измерения общей производительности модели многоклассовой классификации. Он вычисляет средний уровень запоминания по всем классам, придавая каждому классу одинаковую важность без учета классового дисбаланса:

$$UAR = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{A_{ii}}{\sum_{j=1}^{N_c} A_{ij}}, \quad (4)$$

где A – матрица спутанности (*confusion matrix*), N_c – количество классов. Значение *UAR* находится в диапазоне от 0 до 1.

Эксперимент проводился в три этапа:

1) подготовка обучающей выборки;

2) обучение и тестирование классификатора с использованием различных функции ядра и различных размеров длины фрейма анализа речевых сигналов;

3) оценка модели с использованием метрики *UAR*.

Результаты и их обсуждение

На первом этапе проводились эксперименты по подбору числа D извлекаемых МЧКК. Выбирался размер фрейма равный 1024 отчета (~23 мс), классификатор на основе МОВ с линейным ядром и определялся оптимальное число D МЧКК из множества $\{30, 32, 34, 36, 38, 40, 42, 44\}$. Наилучший результат классификации был получен при $D=34$. Это значение затем использовалось в дальнейших исследованиях.

Выполнялись эксперименты с классификатором на основе МОВ и различными ядерными функциями. Известно, что ядерная функция полиномиального ядра имеет параметр deg – степень полинома, а радиальная базисная функция имеет параметр γ . Также сам по себе классификатор на основе МОВ имеет параметр C , который представляет из себя «бюджет» некоторой величины, отражающий степень нарушения зазора решающей границы [9]. Такие нарушения происходят, поскольку данные в признаковом пространстве не могут быть разделены идеально. Таким образом, параметры deg , γ и C являются гиперпараметрами классификатора на основе МОВ. Для выбора оптимальных значений гиперпараметров использовалась техника поиска по сетке (*grid search*). Для deg использовалась сетка значений от 1 до 8 с шагом 1, для параметра γ формировалось 10 значений в диапазоне от $10^{-3,5}$ до $10^{-1,5}$, расположенных эквидистантно в логарифмическом масштабе, для C использовалась сетка из 12 значений от 0,1 до 100, расположенных эквидистантно в логарифмическом масштабе. Поиск оптимальных значений параметров, позволяющих получить наибольшее значение UAR выполнялся для векторов признаков, полученных при различных значениях размера фрейма анализа. Результаты эксперимента приведены в таблице 1.

Таблица 1

Результирующий UAR для классификатора на основе МОВ с различными ядрами			
Размер фрейма	Линейное ядро	Полиномиальное ядро	Радиальная базисная функция ядра
1024	0.458 ($C=0,01$)	0.457 ($C=0,01, \gamma=1, deg=1$)	0.469 ($C=8,11, \gamma=0.0008$)
2048	0.451 ($C=0,1$)	0.45 ($C=0,01, \gamma=1, deg=1$)	0.471 ($C=8,11, \gamma=0.00088$)
4096	0.454 ($C=0,01$)	0.455 ($C=0,05, \gamma=0.1, deg=1$)	0.476 ($C=2.31, \gamma=0.0014$)
8192	0.469 ($C=0,01$)	0.474 ($C=0,05, \gamma=0.1, deg=1$)	0.482 ($C=28.48, \gamma=0.014$)

Наилучшее значение UAR 48,2 % достигается при использовании классификатора с ядром РБФ и МЧКК, рассчитанными на фрейме размером 8192. Поверхность UAR , рассчитанная в ходе поиска по сетке для этой модели, представлена на рис. 5, а. Видно, что более высокое значение параметров C приводит к более гибкому классификатору с более высокой производительностью. При увеличении размера фрейма анализа до 16384 показатель UAR резко падал до значения порядка 0.134.

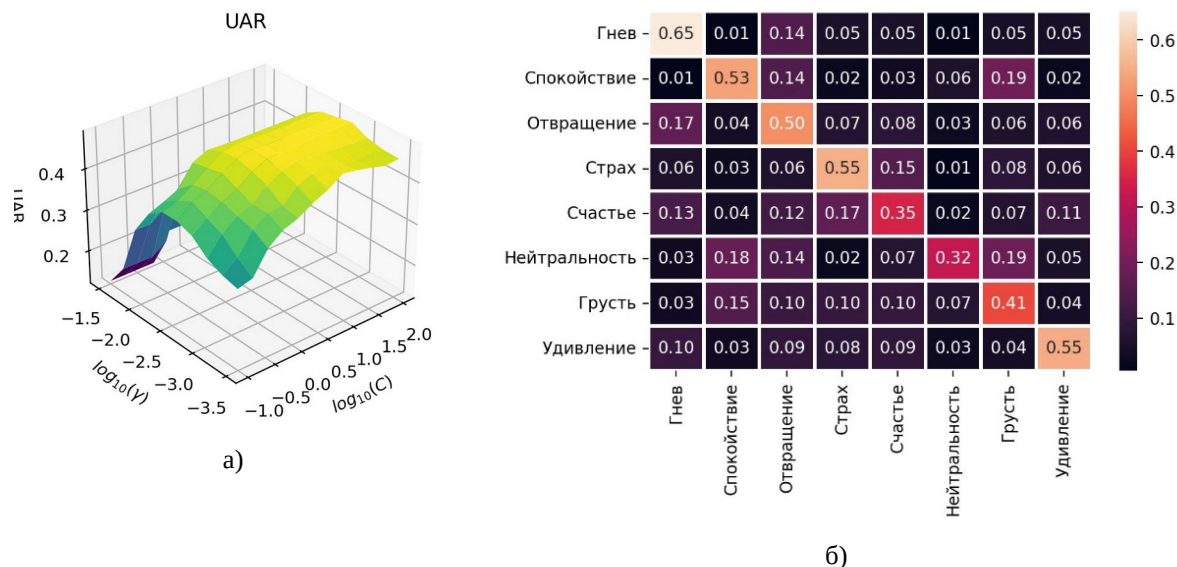


Рис.5. Результаты экспериментов: а) Поверхность UAR; б) матрица спутывания

Fig. 5. а) UAR surface; б) confusion matrix

На рис. 5, б представлена матрица спутывания для лучшей модели SVM-RBF. Анализ матрицы путаницы набора данных RAVDESS с использованием классификатора SVM выявляет важные закономерности в распознавании эмоций. Среди эмоций было замечено, что наиболее часто неправильно классифицированной эмоцией была нейтральность (32%) и счастье (35%). Интересно, что «Нейтральность» часто путают с «Грустью» и «Спокойствием», что позволяет предположить некоторое сходство их акустических характеристик. И наоборот, «Гнев» продемонстрировал высокую точность распознавания (62%) и редко ошибочно классифицировалось как другая эмоция, что указывает на отличительные особенности его акустического профиля. Эти результаты проливают свет на проблемы, с которыми сталкивается классификатор при различении тонких эмоциональных нюансов, и подчеркивают важность разработки функций и совершенствования моделей для улучшения эффективности распознавания эмоций.

Наши результаты показывают, что выбор ядра оказывает существенное влияние на точность классификации. Ядро РБФ продемонстрировало высокую производительность в отношении множества эмоций, в то время как линейное ядро превосходно различало определенные эмоциональные состояния. Примечательно, что размер кадра, используемый для извлечения MFCC, играл значительную роль в общей точности системы: более короткие кадры обеспечивают более мелкие временные детали, а более длинные кадры собирают более широкую контекстную информацию. Эти результаты подчеркивают важность точной настройки ядра классификатора SVM и учета компромиссов, связанных с размером кадра, при разработке систем распознавания эмоций.

Заключение

В сфере взаимодействия человека и компьютера точное распознавание эмоций по речи является ключевым фактором. В этой работе представлен подход к проблеме распознавания речевых эмоций, основанный на классификаторе SVM и сверхсегментарных функциях MFCC. Наилучшие результаты (UAR = 48%) получены при использовании SVM-RBF с характеристиками MFCC, рассчитанными на основе кадров длительностью 185 мс. По сравнению с другими работами [2]–[5] есть возможности для улучшения.

Список литературы / References

1. D. Issa, M. F. Demirci, and A. Yazici, "Speech emotion recognition with deep convolutional neural networks," *Biomedical Signal Processing and Control*, vol. 59, 2020.
2. C. Luna-Jiménez, D. Griol, Z. Callejas, R. Kleinlein, J. M. Montero, and F. Fernández-Martínez, "Multimodal emotion recognition on RAVDESS dataset using transfer learning," *Sensors*, vol. 21, no. 22, pp. 1–29, 2021.
3. Xiao, H.; Li, W.; Zeng, G.; Wu, Y.; Xue, J.; Zhang, J.; Li, C.; Guo, G. On-Road Driver Emotion Recognition Using Facial Expression. *Appl. Sci.* 2022, 12, 807. <https://doi.org/10.3390/app12020807>
4. S. Sadok, S. Leglaive, and R. Séguier, "A vector quantized masked autoencoder for speech emotion recognition," *arXiv preprint arXiv:2304.11117*, 2023.
5. A. Bhavan, P. Chauhan, R. R. Shah et al., "Bagged support vector machines for emotion recognition from speech," *Knowledge-Based Systems*, vol. 184, pp. 1–7, 2019.
6. M. Baruah and B. Banerjee, "Speech emotion recognition via generation using an attention-based variational recurrent neural network," *Proc. Interspeech 2022*, pp. 4710–4714, 2022.
7. Yu, C., Tian, Q., Cheng, F., Zhang, S. (2011). Speech Emotion Recognition Using Support Vector Machines. In: Shen, G., Huang, X. (eds) *Advanced Research on Computer Science and Information Engineering. CSIE 2011. Communications in Computer and Information Science*, vol 152. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21402-8_35
8. X. Huang, A. Acero, H.-W. Hon, and R. Foreword By-Reddy, *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice hall PTR, 2001.
9. T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.
10. C. K. On, P. M. Pandiyan, S. Yaacob, and A. Saudi, "Mel-frequency cepstral coefficient analysis in speech recognition," in *2006 International Conference on Computing & Informatics*, 2006, pp. 1–5.
11. S. R. Livingstone and F. A. Russo, "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 one*, vol. 13, no. 5, p. e0196391, 2018.
12. Luna-Jiménez, C.; Griol, D.; Callejas, Z.; Kleinlein, R.; Montero, J.M.; Fernández-Martínez, F. Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning. *Sensors* 2021, 21, 7665. <https://doi.org/10.3390/s21227665>

Вклад авторов

Вашкевич М.И. осуществлял постановку цели и задачи исследования, а также контроль и проверку решения. Им было предложено использование коэффициента асимметрии, эксцесса и межквантильного размаха для расширения вектора признаков и улучшения точности модели. Разработал стратегию для проведения экспериментов. Принимал участие в подготовке текста статьи и интерпретации результатов экспериментов. Краснопрошин Д.В. выполнил программную реализацию извлечения признаков из датасета (включая расчет МЧКК и их производных, коэффициента асимметрии, эксцесса и межквантильного размаха) и участвовал в подготовке программной базы для эксперимента. Также принимал участие в подготовке текста статьи.

Authors contribution

Vashkevich M.I. carried out setting the goals and objectives of the study, as well as monitoring and verifying the solution to the problem. Proposed to use the skewness coefficient, kurtosis and interquantile range to expand the feature vector and increase model's accuracy. Developed a strategy for conducting experiments. Participated in the preparation of the article as well as the experimental results interpretation. Krasnoproshin D.V. performed a software implementation of extracting features from the initial dataset (including calculation of the MFCC and their derivatives, skewness coefficient, kurtosis and interquantile range) and participated in the preparation of the software base for the experiment. Also took part in the preparation article.

Сведения об авторах

Вашкевич М.И., д.т.н., профессор кафедры

Information about the authors

M.I. Vashkevich Professor, Department of

электронных вычислительных средств (ЭВС)
Белорусского государственного университета
информатики и радиоэлектроники (БГУИР).

Краснопрошин Д.В., магистрант кафедры
электронных вычислительных средств ФКСиС
БГУИР

Electronic Computing Facilities in BSUIR, DrSc.

D.V. Krasnoproshin master student, Department of
Electronic Computing Facilities in BSUIR

Адрес для корреспонденции

220013, Республика Беларусь, г. Минск, ул. П.
Бровки, д. 6, Белорусский государственный
университет информатики и радиоэлектроники
тел. +375-17-293-84-78;
e-mail: vashkevich@bsuir.by
Вашкевич Максим Иосифович

Address for correspondence

220013, Republic of Belarus, Minsk, P. Brovki str.,
6, Belarusian State University of Informatics and
Radioelectronics
tel. +375-17-293-84-78;
e-mail: vashkevich@bsuir.by
Vashkevich Maksim Iosifovich