Эффективное понижения размерности пространства признаков на основе LASSO-регрессии в задачах   
распознавания эмоций по речи

маг. Краснопрошин Д.В., доц. Вашкевич М.И.

Белорусский государственный университет информатики и радиоэлектроники  
ул. П. Бровки, 6, БГУИР, каф. ЭВС, 220013, Минск, Беларусь,   
e-mail: vashkevich@bsuir.by

В докладе показаны различные техники слияния медицинских изображений в области коэффициентов дискретного вейвлет-преобразования (ДВП), полученных как из известных функций, так и на основе целочисленного полифазного представления банка фильтров. Также предложен вариант слияния, использующий представление НЧ-компонент ДВП в области модуля максимума вейвлет-преобразования. Проведено попарное сравнение методов по метрике взаимной информации с применением статистических тестов. Результаты показали, что использование полифазного банка фильтров, а также представление НЧ-компонент ДВП в области модуля максимума вейвлет-преобразования позволяет получить лучше результаты слияния, чем использование двумерного ДВП с функциями Хаара.

# Введение.

**1. КРАТКОЕ введение в проблематику:**

– Обоснование важности распознавания эмоций в речи для различных областей, включая межличностное взаимодействие и технологии искусственного интеллекта.

– Разъяснение актуальности и использования аудиоданных RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) в контексте обучения моделей.

– Поставить акцент на важности задачи понижения признакового пространства. Выделить два аспекта: 1) снижение вычислительной сложности алгоритма распознавания и 2) повышение интерпретируемости модели.

**2**. **Обзор литературы:**

– КРАТКИЙ обзор современных исследований по распознаванию эмоций в речи.

– Оценка существующих методов анализа аудиоданных и выявление их преимуществ и недостатков. – Это достаточно сложная задача в рамках доклада на конференцию, лучше поищите информацию по понижению признакового пространства в контексте задачи распознавания эмоций, если найдете что-то то кратко опишите и сошлитесь.

3. Общая архитектура системы распознавания

– краткое описание

4. Процедура отбора признаков (САМОЕ ГЛАВНОЕ, ЧТО МЫ ХОТИМ ОПИСАТЬ – главный акцент!)

– Сказать, что мы решаем многоклассовую задачу;

– Сведенение K-классовой задачи к K задачам бинарной классификации;

– Описание Lasso для ранжирования признаков, отбор сокращенного числа признаков после ранжирования.

– Объединение признаков, полученных для K бинарных классификаторов.

5. Извлечение признаков:

– КРАТКОЕ описание извлечения признаков, включая мел-кепстральные частотные коэффициенты и их производные.

6. Набор данных RAVDESS:

– КРАТКОЕ описание RAVDESS, включая характеристики аудиоданных, эмоциональные состояния и разнообразие речевых сценариев.

– Обоснование выбора данной базы данных для нашего исследования.

- Внедрение статистических моментов (skewness, kurtosis, interquantile range) в матрицу признаков для более полного охвата характеристик аудиосигнала.

6. Результаты экспериментов

– Представление результатов процесса отбора признаков и их влияния на качество классификации эмоций в речи.

– Подробное рассмотрение ключевых изменений в матрице признаков после применения методов отбора.

– Детальное сопоставление размеров исходной и отобранной матриц признаков.

– Обсуждение результатов в свете сохранения высокой точности классификации при снижении размерности.

– Сопоставление использования различных методов классификации (ЛДА, МОВ, деревья?)

:

7. Заключение

– Краткое подведение итогов и ключевых выводов, основанных на анализе результатов исследования.

– Подчеркивание значимости эффективного отбора признаков для оптимизации методов распознавания эмоций в речи.

– Обсуждение потенциальных направлений для будущих исследований, включая расширение набора данных и оптимизацию методов отбора признаков.

Распознавание эмоций в речи представляет собой важный аспект, оказывающий глубокое влияние на разнообразные сферы нашей жизни, от повседневного межличностного взаимодействия до передовых технологий искусственного интеллекта (ИИ). В современном обществе, где эффективное общение и технологический прогресс играют важную роль, понимание эмоциональных оттенков в речи становится ключевым элементом для создания более гармоничного и интеллектуального окружения.

Эта тема обретает особую актуальность в свете междисциплинарных перспектив, включая медицину, бизнес и технологии, где распознавание эмоций в речи выступает в роли инновационного инструмента, предоставляя новые возможности в области диагностики, маркетинга и даже развития человеко-подобных ИИ-систем.

Для проведения нашего исследования мы использовали набор данных Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [10]. RAVDESS предоставляет уникальную возможность исследования эмоциональных аспектов речи, содержа записи от 24 актеров (12 мужчин и 12 женщин), включая 104 разнообразных высказывания на каждого актера (60 речевых и 44 песенных). В контексте нашей работы, мы ограничились использованием речевых высказываний из RAVDESS, что составило 1440 аудиофайлов в формате wav (16 бит, 48 кГц) с учетом двойной записи каждой вокализации. Этот набор данных обеспечивает широкий спектр эмоциональных состояний, включая нейтральность, спокойствие, счастье, грусть, гнев, страх, удивление и отвращение. Важно подчеркнуть, что эмоциональные состояния были представлены на двух уровнях громкости, что улучшает обучение моделей в условиях повседневного разнообразия эмоций в реальных сценариях общения.

Одним из важных аспектов нашего исследования является задача понижения размерности признакового пространства, что придает проекту значительную актуальность. Это обусловлено не только стремлением к оптимизации вычислительной сложности алгоритма распознавания, что повышает эффективность обработки данных, но и стремлением к улучшению интерпретируемости модели. Понижение размерности признаков позволяет сократить вычислительные ресурсы, улучшая производительность алгоритма распознавания эмоций в речи. Кроме того, это также способствует повышению понимания и интерпретируемости модели, делая ее результаты более понятными и прозрачными для конечных пользователей. Таким образом, снижение размерности признакового пространства становится ключевым фактором в оптимизации алгоритмов, открывая путь к более эффективному и понятному распознаванию эмоций в речи.

Современные исследования в области распознавания эмоций в речи находятся в центре внимания ученых и инженеров, стремящихся улучшить понимание и технологии в этой области. Одним из перспективных направлений является использование глубокого обучения для извлечения высокоуровневых признаков из аудио-данных. Многие исследования фокусируются на применении сверточных нейронных сетей (CNN) и рекуррентных нейронных сетей (RNN), что позволяет более эффективно улавливать временные и частотные закономерности в речевых сигналах.

Параллельно с развитием алгоритмов машинного обучения, активно исследуются методы с использованием активного обучения и обучения с подкреплением. Эти подходы направлены на улучшение производительности систем распознавания, обеспечивая их способность адаптироваться к изменяющимся условиям и повышать точность классификации эмоций в речи.

Одним из вызовов в данной области является разработка универсальных моделей, способных обрабатывать множество языков и культур. Исследователи также активно экспериментируют с комбинированными подходами, включая использование мультимодальных данных, таких как изображения лиц и жесты, для улучшения точности распознавания эмоций в речи.

Также в современных исследованиях по распознаванию эмоций в речи активно изучается применение методов понижения размерности признакового пространства. Это необходимо для того, чтобы оптимизировать вычислительную эффективность и повысить интерпретируемость моделей. Одним из подходов, привлекающих внимание, является использование методов выбора признаков и сжатия данных, таких как анализ главных компонент (PCA) [1]. PCA позволяет выделить наиболее значимые признаки, уменьшая размерность данных и сохраняя при этом максимальное количество информации.Другим подходом является использование вариационного автокодировщика (VAE) [2].

VAE(Auto-Encoding Variational Bayes) предоставляет метод генерации новых признаков, позволяя эффективно понизить размерность признакового пространства и избежать потери информации.Эти методы понижения размерности признакового пространства не только способствуют снижению вычислительной сложности алгоритмов распознавания эмоций в речи, но также значительно улучшают интерпретируемость моделей, делая их результаты более понятными и прозрачными.

Архитектура системы распознавания эмоций в речевом сигнале, рассмотренная в данном исследовании, основывается на использовании мел-частотных кепстральных коэффициентов (МЧКК) и метода опорных векторов (МОВ).

Предложенная модель включает в себя 306-компонентный вектор надсегментных МЧКК признаков в качестве входных данных для классификатора, основанного на методе опорных векторов. Этот вектор представляет собой комплексные характеристики спектра звука, включая информацию о частотах и их интенсивности.

Для оценки качества модели использовалось невзвешенное среднее значение полноты (UAR), что обеспечивает баланс между различными классами эмоций и учитывает несбалансированность данных.

Эксперименты варьировались по нескольким аспектам, включая различные функции ядра для МОВ (линейная, полиномиальная и радиальная базисная функции) и размер фрейма анализа для извлечения МЧКК, который варьировался от 23 до 341 миллисекунд. Это позволяло оценить влияние этих параметров на производительность системы и оптимально настроить ее для достижения максимальной точности распознавания эмоций в речи.

В рамках данной работы особое внимание уделялось отбору признаков. По мнению авторовм это может способствовать повышение эффективности и интерпретируемости модели распознавания эмоций в речевом сигнале. В данном контексте, решая многоклассовую задачу, авторы стремились найти оптимальный подход, который включает в себя следующие важные этапы:

1) Сведение к бинарным классификациям: многоклассовая задача эффективно разбивается на K отдельных бинарных задач классификации, где каждая эмоция рассматривается отдельно. Это позволяет более точно адаптировать модель к каждой эмоциональной категории.

2) Использование МЧКК для извлечения характеристик: МЧКК служит основой для извлечения признаков из речевых сигналов, предоставляя детальный аудио-анализ. Этот шаг формирует входные данные для дальнейшей обработки.

3) Применение метода Lasso для ранжирования признаков: процесс отбора признаков включает в себя использование метода Lasso. Этот механизм не только выделяет наиболее важные характеристики, но и присваивает им веса, что дополнительно улучшает качество модели.

Следует отметить, что **метод Lasso (Least Absolute Shrinkage and Selection Operator)** представляет собой технику регуляризации, которая используется для отбора признаков и снижения размерности данных. Он вводит штраф на сумму абсолютных значений коэффициентов регрессии, что позволяет занулить некоторые из них. В контексте ранжирования признаков, Lasso выделяет наиболее важные характеристики и придает им значимые веса. Алгоритм Lasso минимизирует следующую функцию стоимости:

где:

N - количество наблюдений,

p - количество признаков,

yi - целевая переменная для i-го наблюдения,

xij- значение j-го признака для i-го наблюдения,

βj- коэффициент регрессии для j-го признака,

λ - параметр регуляризации.

Параметр λ регулирует степень регуляризации. При увеличении λ Lasso уменьшает веса признаков, зануляя некоторые из них.

4) Отбор сокращенного числа признаков после ранжирования: после применения Lasso и ранжирования признаков по их важности, следующим шагом является отбор сокращенного числа признаков. Этот этап позволяет выбрать оптимальное подмножество наиболее значимых признаков, снижая размерность данных и устраняя избыточность.

Одним из подходов к отбору сокращенного числа признаков является установка порогового значения на веса признаков и отбор только тех, чьи веса превышают этот порог. Такой механизм позволяет создать более простую и интерпретируемую модель, которая все еще обладает высокой предсказательной силой.

5) Объединение признаков для K бинарных классификаторов: признаки, выделенные для каждой из K бинарных задач классификации, объединяются. Этот этап обеспечивает комплексный и гармоничный подход к распознаванию эмоций в речевом сигнале, где каждая эмоция получает оптимальный вес в итоговой модели.

Согласно предположению авторов такая стратегия отбора признаков может помочь не только обеспечивает точность в решении многоклассовой задачи, но и активно поспособствовать интерпретируемости модели, делая ее результаты более понятными и прозрачными.

На начальном этапе в системе распознавания эмоций по речи выполняется предварительная обработка входных аудиоданных [1, 5]. В данной работе речевые признаки рассчитывались на основании МЧКК [8].

В итоговый набор *исходных признаков* были включены среднее значение МЧКК ( признаков), среднеквадратичное отклонение МЧКК ( признаков), среднее от первой и второй производных от МЧКК (признаков), их среднеквадратическое отклонение (), а также коэффициент асимметрии, эксцесс и межквантильный размах (по признаков для каждой характеристики соответственно). Таким образом, для каждого аудиофайла мы получаем -компонентный вектор надсегментных признаков МЧКК.

Полученный базовый набор признаков далее подавался на вход системе, которая включала в себя метод Lasso для ранжирования признаков и отбора оптимального подмножества характеристик, как было сказано выше.

На рис. 1 показана общая схема применения многомасштабного преобразования для слияния изображений.

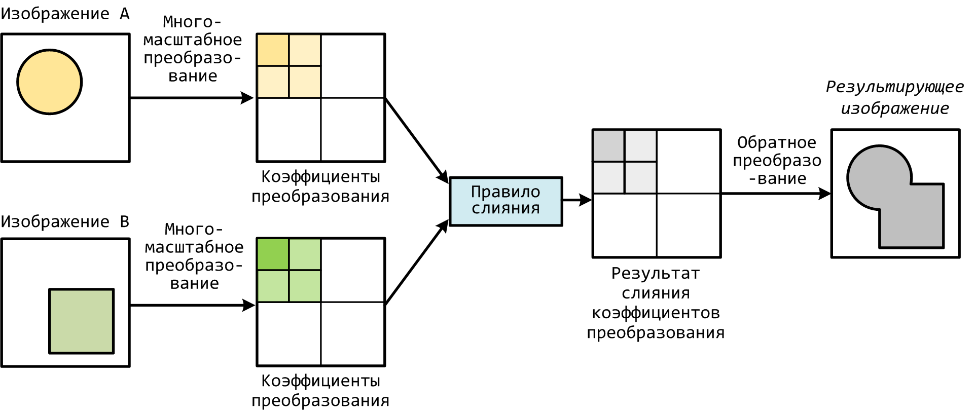


Рисунок 1 – Схема слияния изображения

# Заголовок 1.

Общий принцип построения правил слияния заключается в обеспечении максимальной сохранности всех важных признаков, таких как границы и края. В [5] предложен следующие правила слияния:

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |

где – это соответствующие коэффициенты ДВП изображения .

Согласно (1) НЧ-компоненты, отвечающие за крупные детали объединяются усреднением. Выражение (2) показывает, что слияние ВЧ-коэффициентов ДВП выполняется путем расчета максимума.

# Разреженное представление изображения в пространстве максимумов модуля вейвлет-преобразования.

Результаты экспериментов.

# Заключение.

**Литература**

1. Хлесткин А. Ю. Методы формирования мультимодального изображения //Научные труды SWorld. – 2014. – Т. 4. – №. 4. – С. 14-18.
2. Li S. et al. Pixel-level image fusion: A survey of the state of the art // Information Fusion. – 2017. – vol. 33. – P. 100-112.
3. De Silva D. et al. Wavelet based edge feature enhancement for convolutional neural networks //Eleventh International Conference on Machine Vision (ICMV 2018). – SPIE, 2019. – P. 751-760.
4. Structurally orthogonal finite precision FPGA implementation of block-lifting-based quaternionic paraunitary filter banks for L2L image coding / N.A. Petrovsky, E.V. Rybenkov, A.A. Petrovsky // Digital Signal Processing (DSP’2017): Proc. 22nd Int. Conf., London, UK, Aug. 23-25. – 2017 – 5 P.
5. Qu G., Zhang D., Yan P. Medical image fusion by wavelet transform modulus maxima //Optics Express. – 2001. – Т. 9. – №. 4. – С. 184-190.
6. Дворкович, В.П. Новый подход к использованию вейвлет-фильтров при обработке изображений / В.П. Дворкович, А.В. Гильманшин // Цифровая обработка сигналов – 2008 – № 1 – С. 37–42.
7. Парфенюк М., Петровский А.А. Параунитарные банки фильтров на основе алгебры кватернионов: теория и применение // Цифровая обработка сигналов – 2008 – № 1 – C. 22-36.
8. Qu G., Zhang D., Yan P. Information measure for performance of image fusion //Electronics letters. – 2002. – vol. 38. – №. 7. – P. 313–315.
9. The whole brain atlas [Электронный ресурс]. – 2023. – Режим доступа: <http://www.med.harvard.edu/aanlib/home.html>

FUSION OF MEDICAL IMAGES BASED ON THE   
DISCRETE WAVELET TRANSFORM

Petrovsky N.A., Kuis I.S. Vashkevich M.I.

Belarusian State University of Informatics and Radioelectronics

6, P. Brovki str., Computer Engineering Department, 220113, Minsk, Belarus,  
e-mail: {nick.petrovsky, vashkevich[}@bsuir.by](mailto:}@bsuir.by)

The report shows various techniques for fusion of medical images in the domain of discrete wavelet transform (DWT) coefficients obtained both from Haar functions and based on an integer polyphase representation of the filter bank. A fusion method is also proposed that uses the representation of the low-frequency components of the DWT in the domain of the wavelet modulus maximum. A pairwise comparison of methods based on the metric of mutual information with the use of statistical tests was carried out. The results showed that the use of a polyphase filter bank decomposition, as well as the representation of the low-frequency DWT components in the wavelet modulus maximum domain, allows to obtain better fusion results than using a two-dimensional DWT with Haar functions.