МЕТОД понижения размерности пространства признаков на основе LASSO-регрессии ДЛЯ ЗАДАЧИ   
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: daniil.krasnoproshin@gmail.com, vashkevich@bsuir.by

Доклад посвящен разработке метода понижения размерности пространства признаков, основанного на применении LASSO-регрессии, с целью повышения эффективности распознавания эмоций по речи. В работе рассматривается задача распознавания эмоций, которая остается актуальной в контексте развития систем компьютерного зрения и обработки речи. Предложенный метод представляет собой инновационный подход, который позволяет автоматически отбирать наиболее значимые признаки для классификации эмоций, что способствует улучшению точности и скорости распознавания. В ходе доклада будут рассмотрены теоретические основы метода, его применение к задаче распознавания эмоций по речи, а также результаты экспериментов, подтверждающие эффективность предложенного подхода. Этот исследовательский вклад может быть полезен для разработки более точных и быстрых систем распознавания эмоций, что имеет практическое применение в области компьютерной обработки речи и межличностного взаимодействия.

# Введение.

Распознавание эмоций в речи представляет собой важный аспект, оказывающий влияние на технологии искусственного интеллекта (ИИ).

Одним из перспективных направлений в распознавании эмоций в речи является использование глубокого обучения для извлечения высокоуровневых признаков из аудиоданных. Многие исследования фокусируются на применении сверточных [1] и рекуррентных нейронных сетей [1, 2], что позволяет более эффективно улавливать временные и частотные закономерности в речевых сигналах.

Тем не менее, нейросетевые подходы имеют свои недостатки, которые могут ограничивать их применимость. Одним из основных недостатков нейронных сетей является высокая вычислительная сложность, особенно при обучении глубоких моделей на больших объемах данных. Это может создавать ограничения в доступности ресурсов для исследователей с ограниченным бюджетом или доступом к высокопроизводительным вычислительным системам.

Кроме того, нейронные сети часто характеризуются низкой интерпретируемостью, что затрудняет понимание причинно-следственных связей между входными данными и прогнозами модели. Это усложняет анализ результатов и его понимание, что может быть нежелательным в некоторых приложениях, особенно в медицинской диагностике.

Исследования с использованием более простых подходов остаются актуальными и важными в свете этих недостатков нейронных сетей. Во-первых, они обладают высокой вычислительной эффективностью, что позволяет проводить анализ данных на обычных компьютерах или локальных вычислительных кластерах без необходимости доступа к дорогостоящим вычислительным ресурсам. Во-вторых, они обеспечивают более высокую интерпретируемость результатов, что позволяет исследователям лучше понимать, какие признаки или характеристики влияют на конечный результат классификации.

В контексте использования статистических моделей, таких как метод опорных векторов (МОВ) или линейного дискриминантного анализа, актуальной является задача понижения размерности признакового пространства. Это обусловлено не только стремлением к оптимизации вычислительной сложности алгоритма распознавания, но и стремлением к улучшению интерпретируемости модели. Сокращение размерности характеристического вектора способствует тому, что в результате остаются только признаки имеющие ключевое значение для процесса принятия решения. Все это приводит к интерпретируемости модели, делая ее результаты более понятными и прозрачными для конечного пользователя.

# 1. Разработка системы распознавания эмоций

На рис. 1 представлен процесс разработки системы распознавания эмоций.

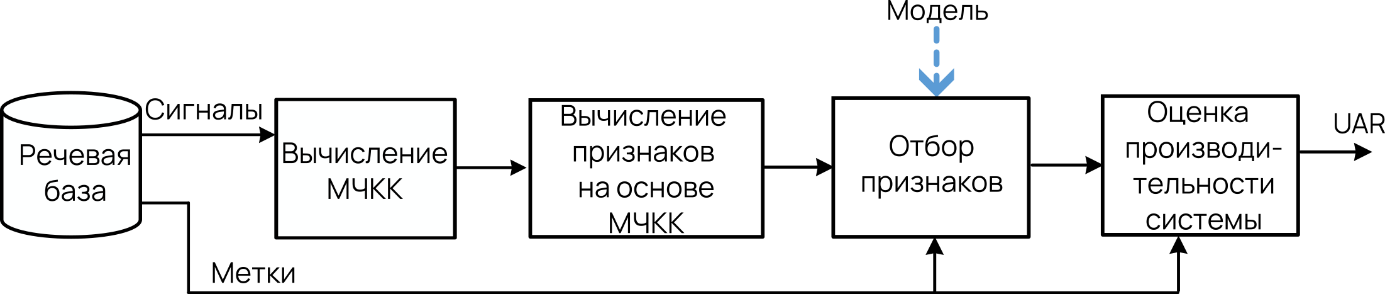


Рисунок 1 – Процесс разработки системы распознавания эмоций по речи

Согласно схеме на рис. 1 процесс разработки основан на использовании аннотированной речевой базы, в которой содержатся образцы речевых сигналов с указанием эмоций с которой они произнесены. Вначале выполняется предварительная обработка аудиосигналов, которая включает вычисление мел-частотных кепстральных коэффициентов (МЧКК), а также их первой и второй производных [3], а также ряда статистик, таких как межквартильный размах, коэффициент ассиметрии и эксцесс. На следующем этапе выполняется отбор признаков, которые способствуют повышению производительности системы распознавания эмоций. Данный этап выполняется с учетом классификатора, который будет использоваться в дальнейшем в системе распознавания эмоций. Отбор признаков позволяет снизить размерность признакового пространства и улучшить обобщающую способность модели, сохраняя при этом высокую точность классификации или даже повышая её. На заключительном этапе производится оценка производительности системы распознавания эмоций с использованием показателя UAR (англ. *unweighted average recall*), вычисляемом при помощи процедуры перекрестной проверки по 5 блокам.

# 2. Процедура отбора признаков

Для отбора признаков разработан ряд методов, таких как LASSO [\*\*6 Джеймс], Relief [5\*\*Флах], mRMR и др. [6\*\* Huang]. Однако, большинство из них разработано в контексте бинарной классификации. В данной работе разработан метод отбора признаков для задачи многоклассовой классификации, основанный на использовании LASSO-регрессии. Метод LASSO (англ. *least absolute shrinkage and selection operator*) представляет собой технику, позволяющую сократить число предикторов в задаче линейной регрессии [\*\*6 Джеймс]. В предыдущих работах [\*\*Tsanas-12, Лихачёв-23] LASSO показал свою эффективность в отборе признаков, применительно к задачам бинарной классификации речевых сигналов. В данной работе предлагается расширить применимость метода LASSO для отбора признаков в задаче многоклассовой классификации. Предлагается следующий алгоритм отбора признаков:

1) Задача классификации на -классов заменяется на задач бинарной классификации, по схеме «один против всех». Это значит, что в начале в один класс относятся все объекты из заносятся в один класс и им присваивается метка “1”, а все остальные объекты заносятся в другой класс и им присваивается метка “0”. Так получается формулируется первая задача бинарной классификации. Затем процедура повторяется, только в отдельный класс помещаются объекты из и т.д. Применительно к рассматриваемой в работе задаче данный этап это означает, что будут получены классификаторы, которые хорошо отделяют одну эмоцию (например, «удивление») от всех остальных.

2) Для решения задач бинарной классификации применяется метод отбора признаков на основе LASSO [Лихачёв-23]. В результате получается подмножеств признаков из исходного полного набора признаков.

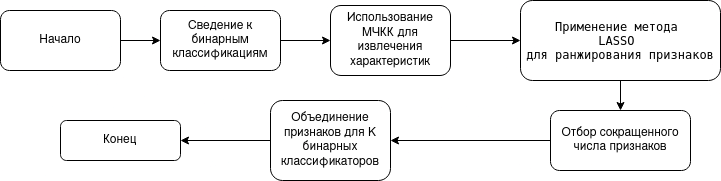
3) Полученные на шаге 2 признаки объединяются для формирования итогового набора признаков.

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.

На рис. 2 представлена схема работы алгоритма:



**Рис. 2.** Схема отбора признаков с помощью алгоритма Lasso

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.

Такой подход позволяет выделить из всего многообразия лишь те признаки, которые способны улучшить производительность системы распознавания эмоций.

**3. Комбинированный метод LASSO для отбора признаков в задаче бинарной классификации.**

В данном разделе более подробно рассмотрен второй этап предлагаемой процедуры отбора признаков.

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.

Метод LASSO (англ. *least absolute shrinkage and selection operator*) [\*\*], используемый в данной работе, основан на решении задачи линейной регрессии:

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

где – число примеров в обучающем наборе, – метка класса -го примера (эмоция), – -й признак -го примера, – -й коэффициент линейной модели, – параметр регуляризации. Решение (1) при достаточно большом приводит к тому, что часть коэффициентов (предикторов) становятся в точности нулевыми. Поэтому решая (1) для ряда возрастающих значений параметра регуляризации  и фиксируя порядок, в котором модель «отбрасывает» признаки можно ранжировать их по значимости (первыми отбрасываются наименее значимые признаки). В этом и заключается метод ранжирования признаков методом LASSO.

В рамках данной работы процесс по отбору признаков в задаче бинарной классификации представляет собой следующее:

1) Подготовка данных: исходные данные включают в себя независимые признаки (X) и целевую переменную (y), которая содержит метки классов (названия эмоций).

2) Переименование и нормализация меток: метки классов переименовываются и нормализуются, чтобы свести их к бинарному виду (-1 и 1, где 1 — целевой класс для данной итерации, а -1 — остальные классы (one vs all).

3) LASSO ранжирование признаков: для каждой эмоциональной метки вычисляется ранжирование признаков с использованием LASSO регрессии. Подбирается оптимальное значение параметра регуляризации с помощью кросс-валидации.

4) Оценка модели с отобранными признаками: для каждой эмоциональной метки производится оценка модели с использованием отобранных признаков. На этом этапе также производится выбор наилучшего подмножества признаков с помощью обратного отбора признаков.

5) Кросс-валидация на выбранном подмножестве признаков: для выбранного подмножества признаков проводится кросс-валидация на данных с учетом разделения по группам.

6) Оценка качества модели: вычисляется усредненная точность по всем фолдам кросс-валидации. После этого строится матрица ошибок для визуализации результатов.

# 4. Извлечение признаков из аудиоданных

В данной работе речевые признаки рассчитывались на основании МЧКК [3]. Расчет МЧКК относится к методам кратковременного анализа речевого сигнала, которые предполагают разбиение сигнала на фреймы (короткие сегменты).

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. Использовагние таких статистических показателей как skewness, kurtosis и interquantile range, а также первой и второй произовных от МЧКК обусловлено стремеление наиболее полно охватить характеристики аудиосигнала.

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.

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

1) Разделить набор разбивается на *k* блоков;

2) В цикле для  выполняются следующие операции:

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

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

– выполняется обучение модели классификатора на тренировочных и оценивоется 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 данных 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 [4]. 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.

***Набор данных RAVDESS.***

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.

Данный датасет обладает контролируемыми условиями записи, разнообразием дикторов и эмоций, а также высоким качеством данных, что делает его надежным инструментом для анализа эмоций в речи.

**5. Модель классификации**

В качестве основного классификатора используется модель на основе линейного дискриминантного анализа.

Линейный дискриминантный анализ (ЛДА) — это статистический метод машинного обучения, который направлен на поиск линейных комбинаций признаков, наиболее эффективно разделяющих данные на различные классы. Принцип работы ЛДА заключается в поиске такого линейного преобразования признаков, которое максимизирует отношение межклассовой дисперсии к внутриклассовой дисперсии. Это достигается путем нахождения оптимальных весовых коэффициентов для линейной комбинации признаков, которые обеспечивают максимальное разделение между классами данных.

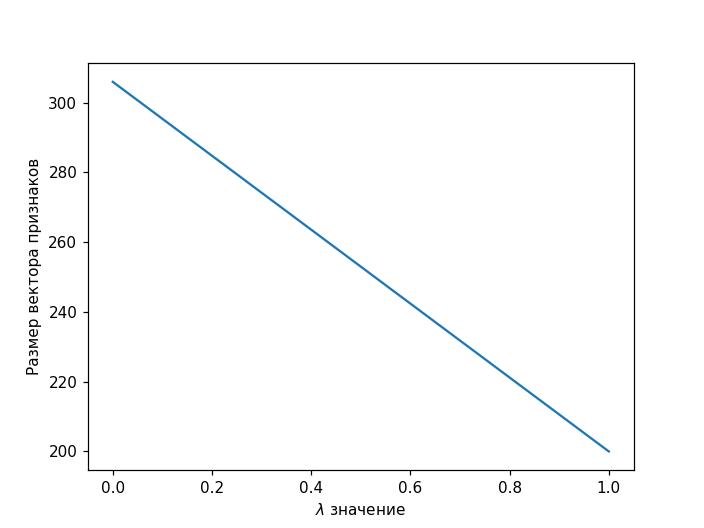
Для дополнительной оценки результатов эксперимента, а также для проведения сравнительного анализа также был использован метод опорных векторов с линейным ядром.

Метод опорных векторов(МОВ) - это алгоритм машинного обучения, который используется для задач классификации и регрессии. Он основан на поиске оптимальной гиперплоскости, которая разделяет данные на два класса. В случае линейного ядра МОВ стремится найти линейную гиперплоскость, которая максимально разделяет классы в пространстве признаков. Это достигается путем максимизации ширины зазора между классами, что обеспечивает лучшую обобщающую способность модели. Таким образом, МОВ с линейным ядром ищет оптимальное разделение данных в исходном признаковом пространстве без необходимости явного преобразования признаков.

**6. Результаты экспериментов**

На первом этапе эксперимента мы провели процедуру переопределения и нормализации данных. Далее, для каждой эмоции мы использовали метод LASSO для ранжирования признаков в порядке их важности для классификации. Для этого мы обучили модель LASSO на различных значениях параметра регуляризации и выбрали оптимальные признаки, которые наилучшим образом коррелируют с каждой из рассматриваемых эмоций.

Дополнительно стоит отметить, что в процессе упорядочивания и оптимизации количества признаков большая часть удаленных характеристик связана с изначально извлеченными мел-кепстральными коэффициентами. В то же время, первая и вторая производные, среднеквадратическое отклонение, а также коэффициенты асимметрии, эксцесса и межквантильный размах, вычисленные на основе исходных мел-кепстральных коэффициентов, почти не были затронуты в процессе отбора признаков. Это свидетельствует о высокой информативности указанных характеристик и их важной роли в распознавании эмоциональных состояний на основе аудиоданных. На рис. 4 проиллюстрирован процесс изменения количества признаков в зависимости от значения лямбды:



**Рис. 4.** Размер вектора признаков в засимости от значения лямбды

Затем мы применили метод анализа дискриминантных признаков (LDA) для оценки классификационной эффективности на основе отобранных признаков. Для каждой эмоции мы последовательно добавляли признаки в модель LDA и оценивали её производительность с помощью метрики UAR (Unweighted Average Recall) на кросс-валидации по оставшимся наблюдениям. Аналогичным образов был применен МОВ с линейным ядром.

Наконец, мы провели анализ ошибок классификации с помощью матрицы спутанности и оценили общую производительность модели на основе метрики UAR.

В итоге, после процедуры отбора признаков из изначального набора, состоящего из 306 характеристик, оставлено 205 признаков при относительном сохранении точности классификатора в случае с ЛДА и 208 признаков в случае с МОВ. Этот результат демонстрирует значительное уменьшение размерности признакового пространства, что является важным шагом в оптимизации аналитических процессов. Одновременно удалось сохранить высокий уровень классификационной точности, что подчеркивает эффективность примененной методологии отбора признаков.

Уменьшение размерности признакового пространства приводит к снижению вычислительной нагрузки и улучшению обобщающей способности классификатора. Это позволяет эффективнее обрабатывать и анализировать данные, что особенно важно в контексте ресурсозатратных задач, таких как анализ эмоций на основе звуковых данных. Более того, сохранение высокой классификационной точности при сокращении числа признаков подчеркивает информативность отобранных характеристик и их значимость для распознавания эмоциональных состояний.

Таким образом, данное исследование демонстрирует успешное сбалансированное сочетание между снижением размерности данных и сохранением качества классификации, что является важным вкладом в развитие методов обработки и анализа данных в области распознавания эмоций.

Этот эксперимент позволяет нам понять, какие аспекты речи наиболее информативны для распознавания различных эмоциональных состояний, что имеет важное значение для развития систем распознавания эмоций на основе звука.

Результаты эксперимента приведены в таблице 1.

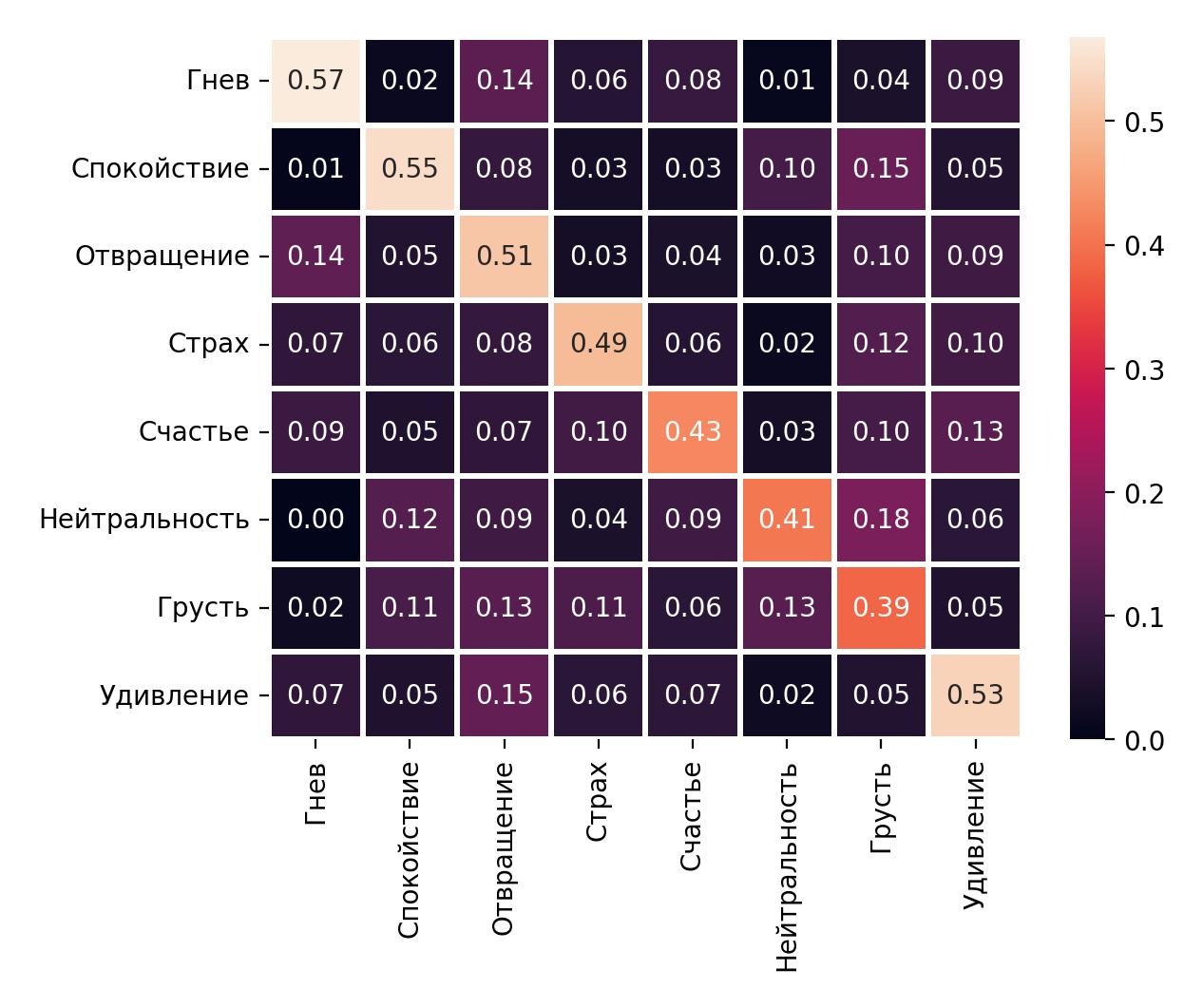
**Таблица 1.** Результирующий UAR для классификаторов на основе ЛДА и МОВ (линейное ядро)

с использованием признаков отобранных с помощью Лассо-ранкинга

**Table 1.** Resulting UAR for the LDA-based and SVM-based (linear kernel) classifiers using features selected via Lasso ranking

|  |  |  |
| --- | --- | --- |
| Классификатор | Полный набора признаков (306) | Набор признаков отобранный с помощью Lasso Ranking |
| ЛДА | 0,460 | **0,484**  **205 признаков** |
| МОВ (линейное ядро) | 0.456 (С = 0.004) | **0.473 (С = 0.008)**  **208 признаков** |

На рис. 5 представлена матрица спутанности для лучшей модели. Анализ матрицы спутанности позволяет выявить важные закономерности в распознавании эмоций. Можно заметить, что наиболее часто неправильно классифицированной эмоцией является грусть (39%) и счастье (41%). Интересно, что «нейтральность» часто путается с «грустью» и «спокойствием», что позволяет предположить некоторое сходство их акустических характеристик. И наоборот, «гнев» имеет высокую точность распознавания (57%) и редко ошибочно классифицировалось как другая эмоция (за исключением отвращения), 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.



**Рис. 5.** Матрица спутывания

**7. Заключение**

Исследование, проведенное на основе анализа Ryerson Audio-Visual Database of Emotional Speech (RAVDES), позволило получить ценные результаты, подтверждающие эффективность методов распознавания эмоций в речи. Основываясь на анализе данных и результатов экспериментов, мы пришли к выводу, что эффективный отбор признаков играет ключевую роль в оптимизации методов распознавания эмоций в речи. Удалось продемонстрировать, что сохранение высокой классификационной точности при сокращении числа признаков позволяет значительно снизить вычислительную нагрузку и улучшить обобщающую способность классификатора.

Подчеркивается важность дальнейших исследований в этой области. В частности, существует потребность в расширении набора данных, что позволит провести более обширные и всесторонние исследования. Также важным направлением является оптимизация методов отбора признаков с целью улучшения производительности классификации и обеспечения более точного распознавания эмоций в речи.

В целом, результаты данного исследования представляют собой значимый вклад в развитие методов распознавания эмоций в речи и указывают на перспективы дальнейших исследований в этой области.

**Литература**

1. Issa D., M. Demirci F., Yazici A. Speech Emotion Recognition with Deep Convolutional Neural Networks. Biomedical Signal Processing and Control, vol. 59, 2020.
2. On C. K., Pandiyan P. M., Yaacob S., and Saudi A. Mel-Frequency Cepstral Coefficient Analysis in Speech Recognition. 2006 International Conference on Computing & Informatics, 2006, pp. 1–5.
3. Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning/ C. Luna-Jiménez, D. Griol, Z. Callejas, R. Kleinlein, J.M. Montero, F. Fernández-Martínez // Sensors. – 2021. – vol. 21. – pp. 1 – 29.
4. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлевают знания из данных / пер. с англ. А.А. Слинкина. М.: ДМК Пресс, 2015. 400 с.
5. Huang S. H. Supervised feature selection: A tutorial //Artif. Intell. Res. – 2015. – Т. 4. – №. 2. – С. 22-37.
6. Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R / Г. Джеймс, Д. Уиттон, Т. Хасти, Р. Тибширани //М.: ДМК Пресс, 2016. – 450 с.
7. Tsanas A. et al. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease //IEEE transactions on biomedical engineering. – 2012. – Т. 59. – №. 5. – P. 1264-1271.
8. Лихачёв Д. С. и др. Комбинированный метод отбора информативных признаков для выявления речевых патологий по голосу //Доклады БГУИР. – 2023. – Т. 21. – №. 4. – С. 110-117.

METHOD for reducing the dimensionality of the feature space based on LASSO regression FOR SPEECH EMOTION RECOGNITION PROBLEM

Krasnoproshin D.V.,Vashkevich M.I.

Belarusian State University of Informatics and Radioelectronics

6, P. Brovki str., Computer Engineering Department, 220113, Minsk, Belarus,  
e-mail: daniil.krasnoproshin@gmail.com, vashkevich[@bsuir.by](mailto:}@bsuir.by)

The report focuses on the development of a dimensionality reduction method based on LASSO regression to enhance the effectiveness of speech emotion recognition. Emotion recognition remains relevant in the advancement of computer vision and speech processing systems. The proposed method presents an innovative approach that automatically selects the most significant features for emotion classification, thus improving recognition accuracy and speed. The presentation covers the theoretical foundations of the method, its application to speech-based emotion recognition tasks, and the experimental results validating its effectiveness. This research contribution can be valuable for the development of more accurate and efficient emotion recognition systems, with practical applications in speech processing and interpersonal interaction.