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 (БГУИР)*  *Кафедра электронных вычислительных средств* |
|  | |

**Цель доклада**

Представить метод понижения размерности пространства признаков, основанный на применении LASSO-регрессии, для повышения эффективности распознавания эмоций по речи.

**Разработка системы распознавания эмоций**

Использовалась аннотированная речевая база с указанием эмоций для каждого образца.

Предобработка данных включала вычисление мел-частотных кепстральных коэффициентов (МЧКК), их первой и второй производных, а также ряда статистик, таких как межквартильный размах, коэффициент ассиметрии и эксцесс.

Затем был произведен отбор признаков, которые способствуют повышению производительности системы распознавания эмоций. Данный этап выполнялся с учетом классификатора, который будет использоваться в дальнейшем в системе распознавания эмоций.

Качество модели оценивалось с помощью UAR(*Unweighted Average Recall*) и перекрестной проверки.

## Процесс разработки системы распознавания эмоций:

|  |
| --- |
|  |

**Процедура отбора признаков**

Разработан алгоритм отбора признаков для задачи многоклассовой классификации, основанный на использовании LASSO-регрессии:

1) Задача классификации на ***K***-классов **{C1,C2,...Ck}** заменяется на ***K*** задач бинарной классификации, по схеме «один против всех».

2) Для решения задач бинарной классификации применяется метод отбора признаков на основе LASSO. В результате получается **k** подмножеств признаков из их исходного полного набора.

3) Для каждого из K наборов признаков, полученных на шаге 2, применяется метод пошагового исключения переменных (англ. *BSS – backward-stepwise selection*). Этот этап позволяет выбрать оптимальное подмножество наиболее значимых признаков, снижая размерность данных и устраняя избыточность.

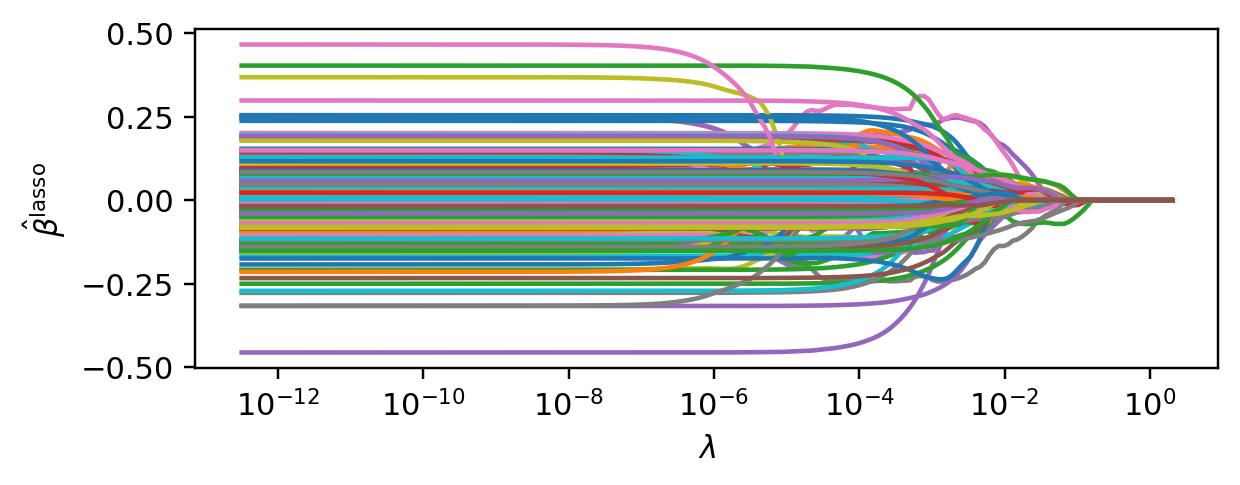
4) Полученные на шаге 3 **K** наборов признаков объединяются для формирования итогового набора признаков.

**Применение LASSO для отбора признаков в задаче бинарной классификации**

Метод LASSO, используемый в данной работе, основан на решении задачи линейной регрессии:

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

где **N**– число примеров в обучающем наборе, **yi** – метка класса **i**-го образца, **xij**– **j**-й признак **i**-го образца, **βj** – **j**-й коэффициент линейной модели, **λ** – параметр регуляризации. Решение (1) при достаточно большом **λ** приводит к тому, что часть коэффициентов (предикторов) **β** становятся в точности нулевыми.



После ранжирования признаков была выполнена оценка полноты классификации при использовании возрастающего набора признаков. В ходе этой процедуры определялось число признаков, которое обеспечивало наибольшую точность классификации.

|  |  |
| --- | --- |
| Отбор подмножества признаков по результатам ранжирования с использованием LASSO-регрессии |  |

**Набор данных и извлечение признаков**

В качестве набор данных использовался Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [3]. который содержит записи от 24 актеров (12 мужчин и 12 женщин), включая 104 высказывания на каждого актера (60 речевых и 44 песенных). В контексте нашей работы, мы ограничились использованием речевых высказываний, что составило 1440 аудиофайлов в формате wav (16 бит, 48 кГц. RAVDESS содержит эмоциональные состояния: нейтральность, спокойствие, счастье, грусть, гнев, страх, удивление и отвращение.

Речевые признаки рассчитывались на основании мел-частотных кепстральных коэффициентов (МЧКК).

В итоговый набор *исходных признаков* были включены среднее значение МЧКК (34 признака), среднеквадратичное отклонение МЧКК (34 признака), среднее от первой и второй производных от МЧКК (68 признаков), их среднеквадратическое отклонение (68), а также коэффициент асимметрии, эксцесс и межквантильный размах (по 34 признака для каждой характеристики соответственно). Таким образом, для каждого аудиофайла мы получаем 306-компонентный вектор надсегментных признаков МЧКК.

Для тестирования классификатора использовался метод перекрестной проверки по *k*-блокам (*k-fold cross-validation*)

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).

**Результаты экспериментальных исследований**

Предложенный метод может применятся с различными моделями классификации, в частности исследовалось применение метода с классификаторами на основе линейного дискриминантного анализа (ЛДА) и метода опорных векторов (МОВ). Оценка производительности системы распознавания эмоций выполнялась с использованием невзвешенной средней полноты – UAR (*Unweighted Average Recall*). Результаты эксперимента приведены в таблице 1.

**Таблица 1.** Результирующий UAR для классификаторов на основе ЛДА и МОВ

|  |  |  |
| --- | --- | --- |
| Классификатор | Полный набора признаков (306) | Отобранный набор признаков |
| ЛДА | 0,460 | **0,484**  (205 признаков) |
| МОВ (линейное ядро) | 0.456 | **0.473**  (208 признаков) |

После процедуры отбора признаков из изначального набора, состоящего из 306 характеристик, оставлено 205 признаков при повышении 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.

Благодаря этому эксперименту нам удалось понять, какие аспекты речи наиболее информативны для распознавания различных эмоциональных состояний, что имеет важное значение для развития систем распознавания эмоций на основе звука.

Далее представлена матрица спутанности для лучшей модели.

|  |  |
| --- | --- |
| **Матрица спутанности** |  |

Проанализировав данную матрицу 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.

Более того, 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.

**Заключение**

В работе *предложен* метод понижения размерности пространства признаков, основанный на применении LASSO-регрессии, для повышения эффективности распознавания эмоций по речи*.*

В работе выполнено сравнение качества класиификации модели с использованием исходного набора признаков, и набора признаков полученных с помощью разработанного метода понижения размерности признакового пространства. Благодаря новому методу **удалось снизить размерность примерно на 30% и одновременно удалось сохранить высокий уровень классификационной точности, что подчеркивает эффективность примененной методологии отбора признаков.**