

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.934.2

Езовит  
Алексей Владимирович

Алгоритм выделения информационных признаков речевого сигнала

**АВТОРЕФЕРАТ**

на соискание степени магистра технических наук  
по специальности 1-40 80 02 «Системный анализ, управление и обработка  
информации»

Научный руководитель  
Гуринович Алевтина Борисовна  
кандидат физико-математических наук,  
доцент

Минск 2023

## ВВЕДЕНИЕ

В современных IT-технологиях вопросам визуальной и акустической обработки данных отводится важное место. У этой тенденции есть ряд причин. Первая заключается в необходимости обработки большого количества информации, собранной с фотографий/аудиозаписей, которую невозможно или крайне сложно получить без применения вычислительной техники. Вторая заключается в возможностях алгоритмов машинного обучения находить в данных информацию и закономерности, не являющиеся очевидными для человека. Еще одна причина заключается в необходимости создания интерфейса для взаимодействия автоматизированных систем (например, голосовых помощников) с окружающим миром и человеком, в частности.

Данное исследование направлено на выделение информационных признаков из множества характеристик, полученных по записи речи. Исследования в данной сфере могут быть полезны в других областях. Например, в человеческой речи можно обнаружить симптомы некоторых болезней. Исследование информационных признаков будет полезно при их диагностировании. В области распознавания эмоционального состояния, пола, возраста, выделение информационных признаков позволит уменьшить затрачиваемые вычислительные мощности, что увеличит эффективность при внедрении в сферу общения между искусственным интеллектом и человеком, расширит возможности в сфере оказания услуг. Исследование характеристик речи может быть применено в психологическом анализе человека, что полезно в медицинской диагностике. Так же выделение из речевого сигнала информационных признаков может помочь составить «портрет» человека, позволяющий определить, кем он является, и при необходимости, идентифицировать его, что может быть полезно в области безопасности и криминалистики.

Целью диссертационной работы является разработка и исследование алгоритма, для выделения из множества входных признаков информационных.

Объект исследований: характеристики, выделяемые из речевого сигнала.

Предмет: процесс выделения информационных признаков из множества характеристик аудиосигнала.

Работа состоит из следующих частей:

1 Обзор и анализ существующих моделей и методов выделения информации из речевого сигнала. В данном разделе приведено исследование научных публикаций по распознаванию различных состояний по аудиосигналу речи. Внимание уделяется работам, направленным на распознавания пола, возраста, эмоционального состояния, симптомов болезней. Проанализированы входные признаки алгоритмов, методы машинного обучения, используемые в сфере обработки аудиозаписей речи, модернизации алгоритмов, предлагаемые в работах;

2 Анализ теоретической базы, которую можно применять для выделения информационных признаков. В данном разделе внимание уделяется специфике органов рече- и голосообразования человека, подходам в

исследуемой области, входным признакам, используемым при распознавании различных состояний (и особенностей) человека, алгоритмам машинного обучения;

3 Разработка алгоритма позволяющего выделить информационные признаки из речи. В данном разделе внимание уделено выбору набора данных, с которым предстоит работать, обработке данных. Предложен алгоритм для выделения и сжатия информационных признаков. Проведено исследование предложенного алгоритма на выбранном наборе данных.

Практическая и научная ценность состоит в том, что предложенный алгоритм позволяет значительно уменьшить количество признаков, обрабатываемых алгоритмами машинного обучения при распознавании различных состояний (и особенностей) говорящего. Это приводит к уменьшению затрачиваемых вычислительных мощностей и позволяет анализировать нюансы аудиосигнала. Данный метод является инновационным в решении данной задачи.

# **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

## **Актуальность исследования**

Выделение информационных признаков из речи человека является важной задачей в области обработки информационного сигнала. Достижения в этой области могут быть применены во множестве областей, таких как медицина, безопасность, маркетинг и др. Разрабатываемый здесь подход позволит сократить потребляемые производительные мощности, установить более четкую связь между множеством характеристик голоса и состоянием, которое они описывает.

## **Цель исследования**

Целью диссертационной работы является разработка и исследование алгоритма, для выделения из множества входных признаков информационных.

## **Задачи исследования**

- 1 Исследование существующих моделей и методов выделения информации из речевого сигнала.
- 2 Анализ теоретической базы, которую можно применять для выделения информационных признаков.
- 3 Разработка и тестирование алгоритма, для выделения из множества входных признаков информационных.

## **Новизна полученных результатов**

Научная новизна заключается в том, что был предложен комбинированный алгоритм выделения информационных признаков из множества характеристик речи, позволяющий уменьшить вектор входных признаков в несколько раз при этом сохраняя, а в некоторых случаях и превышая исходную точность алгоритмов, использующихся для проверки.

## **Личный вклад соискателя.**

Соискателем выполнена работа в полном объеме. Постановка задач и обсуждение результатов проводились совместно с научным руководителем и сотрудниками кафедры информационных технологий автоматизированных систем Белорусского государственного университета информатики и

радиоэлектроники. Соавторы опубликованных работ принимали участие в обсуждении промежуточных и конечных результатов. Обработка, интерпретация данных, а также выводы сделаны автором самостоятельно.

### **Апробация результатов диссертации**

Основные положения диссертационной работы докладывались на следующих научных конференциях:

1. The International Conference on Information Technologies and Systems ITS 2021 (Минск 2021);
2. 58-ой научной конференции аспирантов, магистрантов и студентов;
3. The International Conference on Information Technologies and Systems ITS 2022 (Минск 2022);
4. XIII Международной научно-практической конференции профессорско-преподавательского состава, аспирантов, магистрантов и студентов «Актуальные проблемы правовых, экономических и гуманитарных наук».

## СОДЕРЖАНИЕ РАБОТЫ

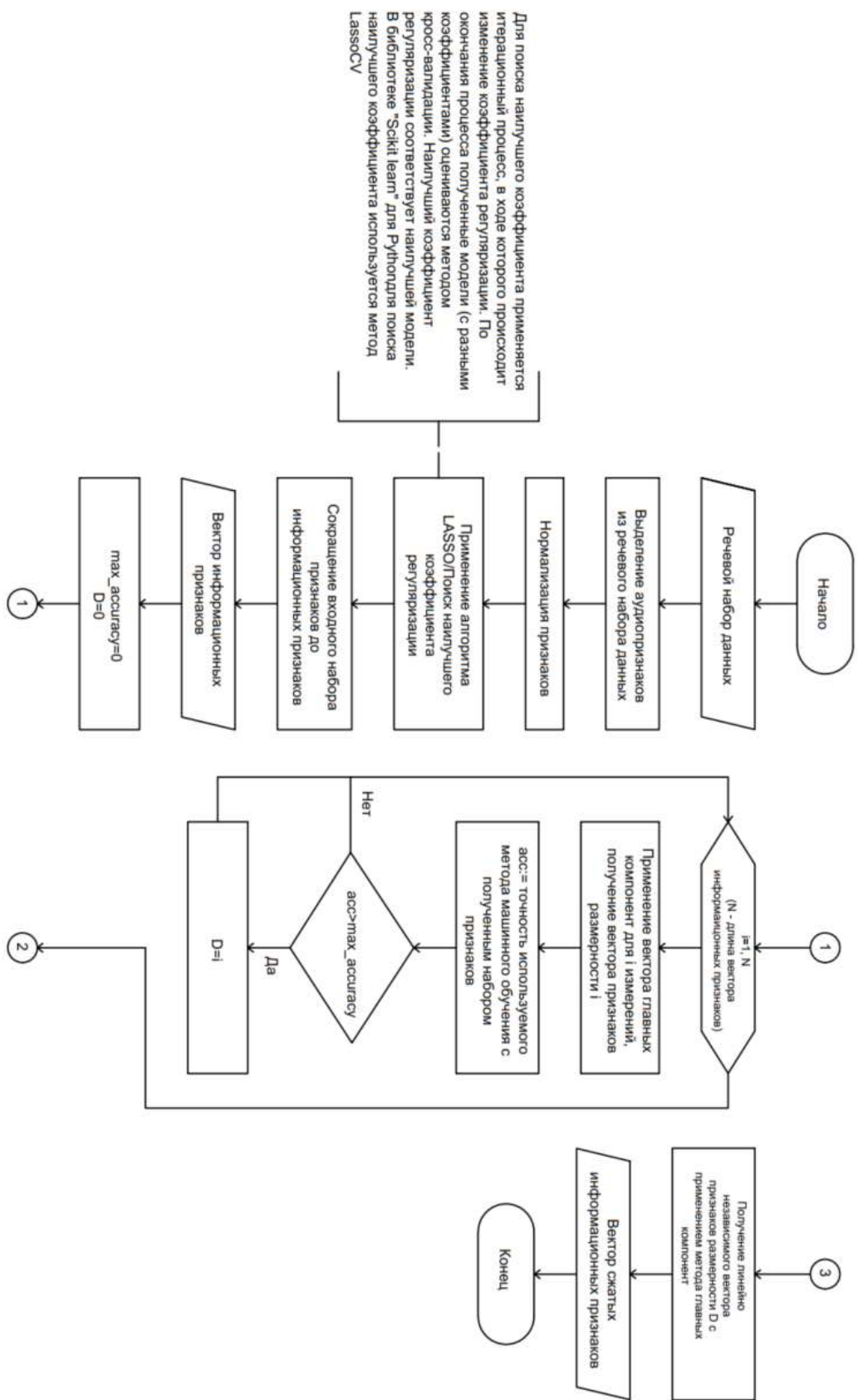
Работа состоит из введения, главы посвященной исследованиям в области обработки речевого сигнала, главы описывающей теорию, лежащую в основе выделения информационных признаков речевого сигнала в данной работе, последней главы, в которой предложен алгоритм для выделения информационных признаков и проведено его исследование и заключения.

В первой главе проведено исследование научных публикаций, в которых авторы предлагают подходы к распознаванию различных состояний человека (эмоциональное состояние, пол/возраст, симптоматику некоторых болезней). По результатам исследований существующих методов была собрана информация о применяющихся методах машинного обучения, используемых входных векторах и достигнутых точностях прогнозирования. В заключении главы сделан вывод, что в исследованиях не уделяется должного внимания анализу входных признаков алгоритмов.

Во второй главе описываются подходы, применяемые в области выделения информационных признаков из речи. Рассматриваются анатомический аппарат речеобразования, характеристики аудиосигнала, использующиеся в обучении моделей и сами модели, показавшие высокую эффективность в данной области.

В третьей главе предложен алгоритм, позволяющий выделить информационные признаки из множества характеристик аудиосигнала и спроецировать их на линейно независимые векторы, дополнительно уменьшая количество признаков, поступающих на вход модели (Рисунок 1).

По результатам исследования было установлено, что исходное признаковое пространство можно сократить до информационного, уменьшая его размерность до 3 раз, при этом потери точности не превышают 4%. При дальнейшем применении метода главных компонент, можно сжать входной вектор еще в 2-4 раза. Потери точности при этом нет. Для некоторых методов удалось достигнуть повышения точности (до 6%).



Для поиска наилучшего коэффициента применяется итерационный процесс, в ходе которого происходит изменение коэффициента регуляризации. По окончании процесса полученные модели (с разными коэффициентами) оцениваются методом кросс-валидации. Наилучший коэффициент регуляризации соответствует наилучшей модели. В библиотеке "Scikit learn" для Python для поиска наилучшего коэффициента используется метод LassoCV

Рисунок 1

## ЗАКЛЮЧЕНИЕ

В ходе исследования был проведен анализ методов и входных векторов алгоритмов, применяющихся в задачах обработки голоса. Было установлено, что алгоритмы обрабатывают векторы признаков избыточной величины. Для решения этой проблемы был предложен комбинированный алгоритм для сокращения входных признаков до информационных и дополнительного уменьшения их размерности (сжатия).

Исследование проводилось на основе задачи распознавания полярности эмоций. Алгоритм применялся три раза для трех обучающих/тренировочных выборок, полученных при разных соотношениях тестовой и тренировочной выборок. Исследование алгоритма проводилось в следующем порядке:

1. Предварительная обработка (выделение признаков, нормализация);
2. Определение наилучшего коэффициента нормализации для алгоритма LASSO;
3. Исследование точности прогнозирования с исходным набором признаков;
4. Сокращение набора входных признаков до информационных;
5. Исследование точности прогнозирования с набором информационных признаков;
6. Сжатие информационных признаков с помощью метода главных компонент;
7. Исследование точности алгоритма с сжатыми информационными признаками для разного количества главных компонент.

Согласно исследованию, было установлено, что наилучшая точность достигается с применением метода опорных векторов и составляет 88% для исходного набора признаков (130), для сокращенного набора признаков точность составляет 87%, а количество признаков сокращается до 53, после сжатия точность составляет 88%, количество признаков 26. При этом уже для количества признаков 8, точность составляет порядка 86%. Для многослойного персептрона исходная точность составляет 84%, точность после сжатия 88% при количестве признаков 24.

Таким образом было показано, что для распознавания полярности эмоций достаточно входного набора признаков низкой размерности (от 3 до 8 в зависимости от размера обучающей выборки). Точность при этом достигает 86 – 87%. В дальнейшем планируется провести исследование по оценке влияния коэффициентов регрессии LASSO, полученных после выделения информационных признаков на точность используемых моделей и возможность их использование в методе главных компонент.



## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Езовит, А. В. Методы выделения информационных признаков речевых сигналов / Езовит А. В., Гуринович А. Б. // Информационные технологии и системы 2021 (ИТС 2021): материалы международной научной конференции – Минск, 2021. – С. 142–143.

2. Езовит, А. В. Подход к выделению информационных признаков речевого сигнала на основе нейросетей / А. В. Езовит // Информационные технологии и управление: материалы 58-ой научной конференции аспирантов, магистрантов и студентов – Минск, 2022. – С. 43.

3. Езовит, А.В. Алгоритм распознавания полярности эмоций /Езовит А.В. (БГУИР), Гуринович А.Б. (БГУИР). // XIII Международной научно-практической конференции профессорско-преподавательского состава, аспирантов, магистрантов и студентов «Актуальные проблемы правовых, экономических и гуманитарных наук» – Минск, 2023.

4. Езовит, А. В. Метод выделения информации из характеристик аудиосигнала / А. В. Езовит, А. Б. Гуринович // Информационные технологии и системы 2022 (ИТС 2022): материалы Международной научной конференции– Минск: БГУИР, 2022. – С. 139–140.