

Федеральное государственное автономное  
образовательное учреждение высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт математики и фундаментальной информатики  
Базовая кафедра вычислительных и информационных технологий

УТВЕРЖДАЮ

Заведующий кафедрой

\_\_\_\_\_ / В.В. Шайдуров

« \_\_\_\_\_ » \_\_\_\_\_ 2020 г.

## **МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**

### **ИССЛЕДОВАНИЕ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ РЕЧИ**

**Направление** 02.04.01 Математика и компьютерные науки

**Магистерская программа** 02.04.01.02 Вычислительная математика

Научный руководитель,

кандидат физико-математических наук,

доцент

\_\_\_\_\_ / И.В. Баранова

Выпускник

\_\_\_\_\_ / С.В. Гилин

Красноярск 2020

## **АННОТАЦИЯ**

Цель работы – исследование алгоритмов распознавания речи, а также разработка собственного алгоритма для решения задачи распознавания слов в аудиозаписях.

Разработан алгоритм гибридной СММ-нейросети, полученный путём смешивания алгоритмов скрытой марковской модели и нейронной сети. Решен практический пример задачи распознавания речи, реализован программный модуль, решающий поставленную задачу, проведен сравнительный анализ разработанного модуля и уже существующих программных обеспечений.

Ключевые слова: Распознавание речи, звук, нейронная сеть, скрытая марковская модель, аудиозапись, синапс, нейрон, обучающая выборка, матрица весов.

## **ANNOTATION**

The purpose of the work is the study of speech recognition algorithms, as well as the development of our own algorithm for solving the problem of word recognition in audio recordings.

An algorithm is developed for a hybrid HMM neural network, obtained by mixing the algorithms of a hidden markov model and a neural network. A practical example of the speech recognition problem is solved, a software module that solves the task is implemented, a comparative analysis of the developed module and existing software is performed.

Keywords: speech recognition, sound, neural network, hidden markov model, audio recording, synapse, neuron, training set, weight matrix.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1   Задача распознавания речи .....	6
1.1 Математическая постановка задачи распознавания образов .....	7
1.2 Задача распознавания речи .....	8
1.3 Особенности задачи и этапы ее решения.....	9
1.4 Звуковые колебания и их представление .....	11
1.5 Мел-кепстральные коэффициенты .....	13
1.6 Методы распознавания речи .....	15
2   Искусственная нейронная сеть.....	18
2.1 Искусственная многослойная нейронная сеть .....	18
2.2 Метод обратного распространения ошибки .....	21
3   Скрытые марковские модели.....	25
4   Гибридная СММ-нейросеть.....	29
5   Практическая задача распознавания.....	32
5.1 Обучающая выборка .....	32
5.2 Требования, предъявляемые к аудиофайлам.....	33
6   Описание программной части .....	34
6.1 Описание использованных библиотек .....	34
6.2 Примеры графического интерфейса и результата работы программного модуля.....	34
6.3 Результаты сравнительного анализа.....	37
ЗАКЛЮЧЕНИЕ .....	40
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	41
ПРИЛОЖЕНИЕ А .....	43

## ВВЕДЕНИЕ

Распознавание речи составляет неотъемлемую часть человеческой деятельности. Разработка систем распознавания речи является одним из самых актуальных и востребованных направлений в области искусственного интеллекта, машинного обучения и анализа данных. Задача распознавание речи относится к классу задач распознавания образов.

На данный момент существует множество частных случаев задачи распознавания образов: оптическое распознавание символов, идентификация людей по лицам, распознавание отпечатков пальцев, распознавание речи, распознавание номеров автомобилей, распознавание объектов ландшафта и другие. В данной работе решается задача распознавания слов в аудиозаписях.

Главный вопрос теории слуха: какие алгоритмы использует мозг для анализа быстро меняющихся звуков окружающего нас мира? Слуховые нейроны сохраняют подробную временную информацию о звуковых особенностях, но мы не знаем, как мозг использует ее для обработки звука.

Распознавание речи применяется в различных областях человеческой деятельности: в голосовом наборе команд, в системах интерактивного речевого самообслуживания, диктовке текстов, в голосовом управлении различной техникой и многих других задачах.

В первой главе диссертации приводятся основные понятия и математическая постановка задачи распознавания образов. Также в этой главе дается классификация методов распознавания образов, и указываются особенности задачи распознавания речи.

Во второй и третьей главах рассматриваются два наиболее популярных подхода к решению задачи распознавания речи: нейронные сети и скрытые марковские модели.

В четвертой главе предлагается алгоритм гибридной СММ-нейросети, представляющий собой объединение подходов, описанных в третьей и четвертой

главе. Разработанный метод позволяет повысить быстродействие распознавания слов в аудиозаписях.

В пятой главе решается практический пример задачи распознавания речи – задачи распознавания слов в аудиозаписях.

Для решения практической задачи и выполнения серии вычислительных экспериментов был создан программный модуль, реализующий работу предложенного алгоритма гибридной СММ-нейросети. В шестой главе описывается разработанный программный модуль, демонстрируется графический интерфейс модуля и особенности его работы. Также выполняется сравнительный анализ разработанного алгоритма распознавания и двух популярных приложений распознавания речи по следующим характеристикам: скорость распознавания, точность распознавания, количество ложных положительных и ложных отрицательных срабатываний (т.н. ошибок первого и второго рода).

#### **Объект исследования.**

В магистерской диссертации объектом исследования выступают методы распознавания речи.

#### **Предмет исследования.**

Предметом исследования является задача распознавания слов в аудиозаписях.

#### **Цель работы.**

Целью магистерской диссертации является разработка и исследование алгоритмов распознавания речи.

Чтобы достичь поставленной цели были поставлены и решены следующие задачи:

- Изучить постановку задачи распознавания образов и основные алгоритмы распознавания;
- Рассмотреть частный случай задачи распознавания – распознавание слов в аудиозаписях;

- Разработать алгоритм решения задачи распознавания речи на основе скрытых марковских моделей и нейронных сетей;
- Создать программное обеспечение, реализующее работу предложенного алгоритма;
- Решить практическую задачу распознавания слов в аудиозаписях;
- Выполнить сравнение разработанного алгоритма распознавания и двух популярных приложений распознавания речи;
- Провести анализ полученных результатов решения практической задачи.

#### **Теоретическая и практическая ценность.**

В магистерской диссертации предложен алгоритм гибридной СММ-нейросети, который представляет собой объединение подходов скрытых марковских моделей и нейронных сетей. Разработанный гибридный алгоритм позволяет сократить время решения задачи и увеличить точность распознавания.

Решение поставленной задачи может быть использовано в облегчении ввода заметок на мобильных телефонах, персональных компьютерах или речевых помощников.

Также система распознавания слов может быть интегрирована в работу искусственных консультантов на web-сайтах. Подобная система позволит сократить время, которое обычно тратится на рукописный ввод.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Воронцов, К. В. Лекции по алгоритмам кластеризации и многомерного шкалирования / К. В. Воронцов. – Москва: МГУ, 2007. – 18 с.
2. Дуда, Р. Распознавание образов и анализ сцен: пер. с англ. Г. Г. Вайнштейна, А. М. Васьковского / Р. Дуда, П. Харт; под ред. В. Л. Стефанюка. – Москва: Мир, 1976. – 502 с.
3. Еремин, Д. М. Искусственные нейронные сети в интеллектуальных системах управления: учеб. пособие / Д. М. Еремин, И. Б. Гарцеев. – Москва: МГИРЭА, 2004. – 34 с.
4. Загоруйко, Н. Г. Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. – Новосибирск: ИМ СО РАН, 1999. – 270 с.
5. Зенин, А. В. Анализ методов распознавания образов / А. В. Зенин // Молодой ученый. – 2017. – №16. – С. 125-130.
6. Лепский, А.Е. Математические методы распознавания образов / А. Е. Лепский, А.Г. Броневиц. – Таганрог: ЮФУ, 2009. – 155 с.
7. Местецкий, Л.М. Математические методы распознавания образов / Л. М. Местецкий. – М.: МГУ, 2004. – 85 с.
8. Патрик, Э. А. Основы теории распознавания образов: пер. с англ. В. М. Баронкина, Б. А. Смиренина, Ю. С. Шинакова под ред. Б. Р. Левина / Э. А. Патрик. – Москва: Советское радио, 1980. – 403 с.
9. Попова, Л. П. Обзор существующих методов распознавания образов. / Л. П. Попова, И. О. Датъев. — М.: Сборник научных трудов, 2007. — 11 с.
10. Курс «Графические модели». – Режим доступа: [http://www.machinelearning.ru/wiki/images/8/83/GM12\\_3.pdf/](http://www.machinelearning.ru/wiki/images/8/83/GM12_3.pdf/)
11. Гилин, С.В. Исследование методов решения задачи распознавания слов в аудиозаписях/ С. В. Гилин // Электронный сборник материалов международной научно-технической конференции студентов, аспирантов и молодых ученых “Перспектив Свободный – 2019”. – Красноярск : СФУ, 2019.

12. Баранова, И. В., Гилин, С. В. Методы решения задачи распознавания слов в аудиозаписях с помощью нейронных сетей и скрытых марковских моделей / И. В. Баранова, С. В. Гилин // Труды XVIII конференции по финансово-актуарной математике и эвентологии многомерной статистики. – Красноярск: СФУ, 2019.

13. Гилин, С.В. Solution of the problem of word recognition in audio records with using of neural networks and Hidden Markov Models / С. В. Гилин // Электронный сборник материалов международной научно-технической конференции студентов, аспирантов и молодых ученых “Перспектив Свободный – 2020”. – Красноярск : СФУ, 2020.

14. Баранова, И. В., Гилин, С. В. Solution of the problem of word recognition in audio records with using of neural networks and Hidden Markov Models / И. В. Баранова, С. В. Гилин // Труды XIX конференции по финансово-актуарной математике и эвентологии многомерной статистики. – Красноярск: СФУ, 2020.



Федеральное государственное автономное  
образовательное учреждение высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт математики и фундаментальной информатики  
Базовая кафедра вычислительных и информационных технологий

УТВЕРЖДАЮ

Заведующий кафедрой

\_\_\_\_\_ / В.В. Шайдуров

« \_\_\_\_\_ » \_\_\_\_\_ 2020 г.

## МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

### ИССЛЕДОВАНИЕ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ РЕЧИ

**Направление** 02.04.01 Математика и компьютерные науки

**Магистерская программа** 02.04.01.02 Вычислительная математика

Научный руководитель,

кандидат физико-математических наук,

доцент

*И.В. Баранова* / И.В. Баранова  
22.06.2020

Выпускник

*С.В. Гилин*, 22.06.20 / С.В. Гилин

Красноярск 2020