

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

УДК _____

Яценко
Дмитрий Валерьевич

Распознавание речи по движению мышц лица с использованием алгоритмов
машинного обучения

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1–31 80 10 «Теоретические основы информатики»

Научный руководитель
Ивашенко Валерьян Петрович
кандидат технических наук

Минск 2018

Краткое введение

Чтение по губам — это задача распознавания текста по движению губ. Развитие компьютерной техники ведет к совершенствованию интерфейсов между человеком и компьютером. Один из важнейших способов человеческой коммуникации — речь, поэтому способность распознавания речи является важным направлением совершенствования человеко-машинных интерфейсов. Визуальное распознавание речи может применяться в следующих целях:

- распознавание речи двух и более человек;
- биометрическая идентификация;
- разработка улучшенных слуховых аппаратов;
- управление интеллектуальными системами.

Согласно последним исследованиям, точность распознавания текста по губам человека возрастает для более длинных слов, что указывает на важность функций, фиксирующих временной контекст при неопределенном канале связи.

Разрабатываемая модель преобразовывает последовательность видеофрагментов с переменной длиной слова в текст, используя пространственно-временные свертки, рекуррентную нейронную сеть и функцию потерь нейросетевой временной классификации.

Общая характеристика работы

В ходе работы будет разработана интеллектуальная система, позволяющая распознавать речь человека на русском языке по движению мышц лица и губ в видеопотоке. Для разработки системы будет использована глубокая рекуррентная нейронная сеть. При обучении нейронной сети будет использован метод нейросетевой темпоральной классификации.

На входе в нейронную сеть подается последовательность кадров, для каждого кадра определяется положение губ. Информация пропускается через нейронную сеть, которая на выходе выдает произнесённое слово. Для обучения нейронной сети будет создан набор видео данных.

Задача проекта разработать систему распознавания речи по движению мышц лица с использованием алгоритмов машинного обучения.

Актуальность темы магистерской диссертации: Разработанная система позволит с высокой точностью осуществлять следующие функции:

- Распознавание речи в условиях зашумлённости;
- Распознавание речи для лиц с частичной или полной потерей слуха;
- Биометрическая идентификация;
- Дистанционное управление интеллектуальными системами;
- Системы скрытой передачи информации.

Задача исследования:

- 1) исследование технологий в области машинного зрения и алгоритмов машинного обучения;
- 2) проектирование структуры и архитектуры системы распознавания речи;
- 3) разработка интеллектуальной системы;
- 4) создание выборки данных и обучение нейронной сети;
- 5) тестирование системы.

Объект исследования: Коммуникационный аппарат человека в рамках человеко-машинной системы речевой коммуникации.

Предмет исследования: методы распознавания речи на основе нейронных сетей.

Текст обоснования

В ходе работы будет разработана интеллектуальная система, позволяющая распознавать речь человека на русском языке по движению мышц лица и губ в видеопотоке. Для разработки системы будет использована глубокая рекуррентная нейронная сеть. При обучении нейронной сети будет использован метод нейросетевой темпоральной классификации.

На входе в нейронную сеть подается последовательность кадров, для каждого кадра определяется положение губ. Информация пропускается через

нейронную сеть, которая на выходе выдает произнесённое слово. Для обучения нейронной сети будет создан набор видео данных.

Краткое содержание работы

В первой главе описываются теоретические аспекты создания интеллектуальной системы распознавания речи.

Вторая глава содержит описание проектирования архитектуры системы, количество слоёв и параметров модели нейронной сети. Изменения в архитектуру для поддержки русского языка. Приведены обоснования выбора типа нейронной сети, алгоритма классификации, способа обучения.

В третьей главе описана структура программных компонентов, разработанных для работы модели с русским языком.

Четвертая глава описывается процесс создания набора данных из 1500 видеофрагментов для 3-х чтецов для обучения.

Пятая глава содержит информацию о процессе обучения, скорость, количество эпох. Так же проведён анализ каждого набора данных и изображены графики функции потери показатели ошибки CER и WER.

В главе 6 показана работа нейронной сети на отдельных видеофрагментах.

Заключение

В ходе данной работы была разработана система распознавания речи по движению мышц лица с использованием алгоритмов машинного обучения. Данная система позволяет распознавать русскую речь по движению мышц лица и губ в видеопотоке. Для разработки системы была использована глубокая рекуррентная нейронная сеть. При обучении нейронной сети был использован метод нейросетевой темпоральной классификации.

Были решены следующие задачи:

1) Обзор существующих технологий и алгоритмов распознавания образов в системе человек-машина. Анализ её составных элементов ориентированных на процессы речевого взаимодействия;

Был осуществлен обзор технологий и алгоритмов распознавания образов в системе человек-машина, проанализированы основные элементы системы человек-машина.

2) Проектирование структуры и архитектуры системы машинного чтения по губам;

В ходе работы была спроектирована структура и архитектура системы машинного чтения по губам, подобраны гиперпараметры слоев сети.

3) Разработка интеллектуальной системы машинного чтения по губам;

В ходе работы были разработаны программные компоненты для работы системы чтения по губам для русского языка, а именно модель нейронной сети, модель видеофрагмента, реализация функции вычисления ошибок CER и WER и другие.

4) Создание набора данных для обучения и тестирования системы машинного чтения по губам;

В ходе работы был создан набор данных из 1500 видеофрагментов, составленных 3-я чтецами. Время, затраченное на создание набора данных 40.5 часов. Модель была обучена с помощью данного набора данных, мини-группами по 8 видеофрагментов. Всего было проведено 202 эпохи обучения за 50.5 часов.

5) Обучение и тестирование системы машинного чтения по губам.

Разработанная система чтения по губам для русского языка позволяет с точностью CER=0.17 и WER=0.36 осуществлять распознавание речи человека по движению мышц лица. Точность аналога для английского языка LipNet составляет CER=0.024 и WER=0.066. Дальнейшее повышение точности системы возможно путем увеличения набора данных для обучения.

Оригинальность работы в системе «www.antiplagiat.ru» составляет 82,47%.

Список опубликованных работ

1 – А. Яценко Д.В., Ивашенко В.П. Система машинного чтения по губам с использованием алгоритмов глубокого обучения // BigData and Advanced Analytics 2018, 6 международная научно-практическая конференция, 3 – 4 Мая 2018 / Минск, Беларусь – 2018. – с.236–243.

2 – А. Яценко Д.В. Система машинного чтения по губам с использованием алгоритмов глубокого обучения // 54-я научно-техническая конференция аспирантов, магистрантов и студентов БГУИР, 23 – 27 Апреля 2018 / Минск, Беларусь – 2018.