

ПОДХОД К ВЫДЕЛЕНИЮ ИНФОРМАЦИОННЫХ ПРИЗНАКОВ РЕЧЕВОГО СИГНАЛА НА ОСНОВЕ НЕЙРОСЕТЕЙ

Рассматривается подход в выделение информационных признаков из речевого сигнала на основе нейросетей. Внимание уделяется сверточным и рекуррентным нейронным сетям.

ВВЕДЕНИЕ

Главной целью выделения информационных признаков является сбор информации об ораторе, такой как пол, возраст, эмоциональные черты и др. Данная задача не получила большого распространения на сегодняшний день, наиболее близкой к ней является задача идентификации говорящего. Для ее решения применяются методы, основанные на метрике, оценке вероятности или нейронных сетях. Данная работа рассматривает подходы, основанные на использовании сверточных (Convolutional Neural Networks – CNN) и рекуррентных нейронных сетей (Recurrent Neural Networks – RNN).

I. Анализ существующих подходов

На сегодняшний день для выделения информационных признаков речевого сигнала существуют подходы, основанные на классических алгоритмах и нейронных сетях. В основе методов, основанных на классических алгоритмах, лежит формирование групп из выборки уже имеющихся голосов формирующих определенный интервал, после чего проводится проверка имеющегося образца на входимость в этот интервал. Тем не менее, из-за роста вычислительных мощностей на сегодня большую популярность получили методы, базирующиеся на глубоком обучении. Согласно статистике 2015 года, точность классических алгоритмов составляла порядка 75%, нейронных сетей – 96%. Однако достижение высокой точности становится возможным в результате наличия большого количества обучающих примеров.

II. Предлагаемые подходы

Предлагаемые подходы рассматривают использование CNN и RNN для выделения информационных признаков.

Согласно исследованию, сверточные нейронные сети стали основным инструментом в области исследования изображения и речи [1]. CNN решают три следующие задачи: классификации (определяет, что представлено на изображении),

детекции (находит область изображения, ответственную за тот или иной класс), сегментации (определяет, какой пиксель изображения к какому классу относится) и наиболее часто применяются при работе с графикой. При подходе основанном на CNN, необходимо представить речь в ее спектральном представлении – спектрограмме. Суть данного подхода сводится к предположению того, что спектрограмма содержит в себе большое количество информации, включая информационные признаки речи. Реализация сводится к следующему: на вход сверточной нейронной сети подается набор спектрограмм с заранее известными признаками, по данному набору сеть учится находить элементы изображения, ответственные за тот или иной признак.

Рекуррентные сети, являются основным методом в исследовании естественного языка. Главным отличием рекуррентных сетей от остальных является наличие циклических связей в скрытом слое. В задачах выделения информационных признаков, подход основанный на RNN кардинально отличается от других подходов. Его суть сводится не к обработке речевого сигнала, а к обработке смысловой информации, содержащейся в речи. Цель RNN в данной задаче заключается в обнаружении культурно-социальных и грамматических конструкций и высказываний, содержащихся в речи и свойственных той или иной группе населения.

III. Выводы

Предлагаемые подходы, основанные на нейронных сетях, позволят выделять из речи человека информационные признаки, содержащие большое количество информации о говорящем. В дальнейшем данную информацию можно использовать для различных задач, например: идентификации человека, синтезе речи, определении состояния человека и др.

1. MDPI[Электронный ресурс]. – Базель, 2021 – Режим доступа: <https://www.mdpi.com/2076-3417/11/8/3603/htm> – Дата доступа: 07.04.2022.

Езовит Алексей Владимирович, магистрант кафедры информационных технологий автоматизированных систем БГУИР, ezovit99@yandex.ru.

Научный руководитель: Гуринович Алевтина Борисовна, канд. физ.-мат. наук, доцент, зам. декана ФИТиУ, БГУИР, gurinovich@bsuir.by.