

МЕТОДЫ РАЗРАБОТКИ СИСТЕМ РАСПОЗНАВАНИЯ РЕЧИ ДЛЯ РАБОТНИКОВ ИТ ИНДУСТРИИ

Медведев Е.Д. (Университет ИТМО), Меженин А.В. (Университет ИТМО)

Научный руководитель – к.т.н., Меженин А.В. (Университет ИТМО)

В данной работе рассматривается проблема распознавания речи при разговоре на темы, связанные с информационными технологиями. Приведены результаты эксперимента по качеству распознавания аудио с речью, содержащей компьютерный сленг, профессионализмы и термины из сферы информационных технологий, и сравнение данного показателя качества со средним значением качества при распознавании речи на общие тематики. Был рассмотрен принцип работы системы автоматизированного распознавания речи, ключевые элементы и метод влияния на качество распознавания. Помимо этого, для эксперимента подготовлен набор данных, содержащий аудио с речью на тематику информационных технологий. Представлен процесс подготовки данного датасета и его состав. Проведена адаптация большой языковой модели от компании Aiphacefei и сравнение результатов качества распознавания модифицированной языковой модели и исходной.

Введение. С точки зрения применения технологий машинного обучения и анализа мультимедийных данных задача обработки естественного языка занимает отдельное место. В современном мире, стремящемся получить, обработать и использовать как можно больше информации, данная задача начинает получать все большую популярность. Сегодня, в повседневную жизнь все активнее внедряются голосовые ассистенты, такие как Алиса, Siri или Google Assistant, для того чтобы упростить нашу жизнь и получить новые данные, которые можно обработать и использовать. Самой главной задачей, которую необходимо решить, для разработки таких помощников, является распознавание речи и перевод ее в понятный для машины или человека вид.

В связи с пандемией COVID-19 и переводом большинства сотрудников на удаленную работу все большее количество компаний задумывается о внедрении автоматизированной системы распознавания речи (ASR) во внутренние процессы компании, с целью упростить процесс составления отчетов и протоколов проведения внутренних собраний. Ввиду различной специфики работы компаний, а также, в большинстве своем, передачи данных на сервера другой компании для распознавания речи, использование готовых решений из коробки не является решением проблемы.

Основная часть. Наиболее популярные системы распознавания русской речи позволяют качественно распознавать спонтанную речь на общие тематики и широко применяются в повседневной жизни. Но ввиду широкого спектра тем для распознавания, подобные системы гораздо хуже справляются с распознаванием речи содержащие в себе термины, профессионализмы и различные словообразования специфичные для сферы информационных технологий.

В данной работе рассматривается проблема распознавания речи людей, употребляющих профессионализмы, компьютерный сленг и термины, связанные с информационными технологиями, на базе системы распознавания речи с открытым исходным кодом – Vosk, ввиду наилучшего качества распознавания среди open-source решений, и большой предобученной языковой моделью от компании Aiphacefei. Рассматриваются особенности речи на темы, связанные с информационными технологиями, и вырабатываются требования к обучающим данным для улучшения показателя качества распознавания. В рамках работы, на тестовых данных проведен эксперимент для определения качества распознавания вышеупомянутой речи с использованием исходной языковой модели и модифицированной, и сравнение полученных результатов.

Выводы. Модифицированная языковая модель имеет преимущество над исходной, поскольку имеет адаптированные под специфическую речь параметры, в частности, перенастроенные веса вероятностей появления слов в речи и обновленный список словоформ. Вследствие чего, модифицированная языковая модель более качественно распознает речь, содержащую различные профессионализмы, термины и сленг, специфичный для специалистов индустрии информационных технологий.

Медведев Е.Д. (автор)

Подпись

Меженин А.В. (научный руководитель)

Подпись