

УДК 004.934.1'1

Исследование регуляризации VPE-dropout в обучении интегральных систем автоматического распознавания речи для турецкого языка

Подлужный И.А., Андрусенко А.Ю., Лаптев А.А.

(Федеральное государственное автономное образовательное учреждение высшего образования “Национальный исследовательский университет ИТМО”)

Научный руководитель – к.т.н., с.н.с. корпоративной лаборатории технологий человеко-машинного взаимодействия Меденников И.П.

(Федеральное государственное автономное образовательное учреждение высшего образования “Национальный исследовательский университет ИТМО”)

В данном докладе исследуется методика обучения интегральных систем распознавания речи для турецкого языка. В рамках рассматриваемой методики также исследуется влияние метода аугментации и регуляризации обучения VPE-dropout на качество обучаемой системы.

Введение: Интегральные системы распознавания речи показывают наилучшие результаты для популярных языков. Однако, для их обучения необходимы большие объёмы данных, порядка сотен часов и больше. Таким образом, обучение интегральных систем распознавания речи в случае, когда обучающих данных на целевом языке недостаточно – малоресурсных языках – представляет особенный интерес. Одним таких малоресурсных языков является турецкий язык. Другой важной задачей систем распознавания речи в речевых ассистентах является распознавание слов из пользовательского словаря, таких как имена, названия организаций и т.п., которых нет в обучающих данных (OOV – out-of-vocabulary слова). В данной работе рассматривается метод обучения интегральной системы распознавания турецкой речи на наборе данных Babel Turkish (80 часов спонтанной речи) с большим содержанием OOV слов (6%).

Основная часть: Для решения проблем обучения систем на малоресурсных языках и распознавания OOV слов используется техника динамической аугментации и регуляризации, основанная на VPE-dropout. VPE-dropout рандомизировано токенизирует обучающие данные увеличивая долю токенов из 1-2 символов и увеличивая число контекстов в котором токены встречаются, что даёт более независимую информацию о произнесениях токенов для обучения системы. В итоге, обучение с применением VPE-dropout регуляризации для рассматриваемой интегральной системы даёт устойчивое улучшение качества распознавания как минимум на 6% относительного WER (word error rate — процент ошибки в словах) и увеличивает долю распознаваемых OOV слов на 25% относительного F-score без привлечения дополнительных вычислительных мощностей. Разработанная система распознавания речи показывает качество 22% CER (character error rate – процент ошибки в символах) процент ошибок в символах) и 38.9% WER на выбранном наборе данных.

Выводы: Представленный метод позволяет эффективно обучать интегральные системы распознавания речи для турецкого языка. В ходе исследования также описано влияние VPE-dropout на обучение таких систем и на их результаты распознавания.

Подлужный И.А. (автор)

Меденников И.П. (научный руководитель)