

Двухязычное русско-казахское распознавание речи с переключением кодов

Д. С. Убский

(Университет ИТМО, г. Санкт-Петербург)

Научный руководитель – д. т. н., Ю. Н. Матвеев

(Университет ИТМО, г. Санкт-Петербург)

Исследования выполнены за счет стартового финансирования университета ИТМО в рамках НИР № 618278 «Синтез эмоциональной речи на основе генеративных составительных сетей»

В данной работе выполнено обучение модели распознавания речи на русском и казахском языках. Архитектурой нейронной сети была выбрана 3-слойная двунаправленная рекуррентная сеть с долгой кратковременной памятью (BLSTM).

Базовые модели BLSTM показали вероятность пословной ошибки (WER) 52.38% на казахском и 31.91% на русском языке. В качестве признаков были использованы значения узкого горлышка многоязычной модели [1], обученной на данных Babel [3].

Модель двухязычного распознавания научилась предсказывать общий набор графем русского и казахского языков [2]. Первые двухязычные модели дали значимый прирост 3% WER на произнесениях на казахском языке за счёт ухудшения WER на 6% на русском языке. Дальнейшие улучшения достигались изменением размерности сети и перестраиванием связывания скрытых состояний НММ с учётом новой разметки данных.

В процессе настройки параметров модели удалось добиться прироста WER на 10% на казахском языке и 0.8% на русском языке (52.38% → 42.22% и 31.91 → 31.12% соответственно).

Литература

1. Y. Khokhlov, I. Medennikov, A. Romanenko, V. Mendeleev, M. Korenevsky, A. Prudnikov, N. Tomashenko, and A. Zatvornitsky, “The STC keyword search system for OpenKWS 2016 evaluation.” in Interspeech, 2017.
2. Suyoun Kim, Michael L. Seltzer, Towards Language-Universal End-to-End Speech Recognition, arXiv: 1711.02207
3. IARPA Babel program [Электронный ресурс]. – Режим доступа: <https://www.iarpa.gov/index.php/research-programs/babel>