

УДК 81'322.6

АНАЛИЗ СОВРЕМЕННЫХ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ СИНТЕЗА РЕЧИ

Абросимов К.И. (Университет ИТМО)

Научный руководитель – к.ф.-м.н. Рыбин С.В. (Университет ИТМО)

В работе рассматриваются различные архитектуры нейронных сетей для задачи синтеза речи. Применяются в данной области как рекуррентные нейронные сети, так и различные механизмы внимания для решения проблем, связанных с надежностью, скоростью и так далее. Особое внимание в работе уделяется характерным нейросетевым механизмам для каждого направления в синтезе речи.

Методы синтеза речи активно внедряются в нашу жизнь, автоматизируя множество рутинных процессов. Синтез речи развивается еще с прошлого столетия, но основные результаты были получены лишь несколько лет назад, благодаря развитию такой области машинного обучения, как глубокое обучение.

Одной из первых нейросетевых моделей считается Tacotron – он состоит из нескольких рекуррентных сетей, однако использует один из самых простых вокодеров – Griffin-Lim. Однако уже в Tacotron 2 уменьшили количество LSTM-слоев и исправили недочет с вокодером, применив нейросетевую вокодер – WaveNet. Основными направлениями в синтезе речи являются – синтез выразительной речи (механизмы Style Tokens), надежный синтез речи (модели NAT), быстрый синтез речи для использования в реальном времени (модели семейства FastSpeech), синтез речи для малоресурсных языков (механизмы Self-supervised Training, Cross-lingual Transfer), адаптивный синтез речи для моделирования определенного голоса (Cross-Acoustic Adaptation, Zero-Shot Adaptation). Каждая из этих тем очень важна для бизнеса и в крупных компаниях, научно-исследовательских центрах, университетах активно развивается.

Анализ современных нейросетевых моделей позволил углубиться в область синтеза речи и направлений исследований в нём для дальнейшего применения их в жизни с помощью голосовых роботов, голосовых ассистентов, технологий человеко-машинного взаимодействия.