

ПОДХОДЫ К ОБУЧЕНИЮ ПРЕДСТАВЛЕНИЙ ДЛЯ ИНТОНАЦИИ В СИСТЕМАХ СИНТЕЗА РЕЧИ НА ОСНОВЕ DNN

Коростик Р.Е. (Университет ИТМО)

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

В докладе рассматриваются последние работы по обучению представлений для интонации в нейросетевых системах синтеза речи, а также результаты воспроизведения некоторых из них.

Современные подходы к построению систем синтеза речи основываются на методах глубинного обучения. При успешном ходе процесса стохастической оптимизации обученная нейронная сеть (при использовании базовой архитектуры) способна генерировать достаточно естественную речь; для коротких высказываний такая речь по естественности вплотную приближается к человеческой.

Тем не менее, один и тот же текст можно прочитать по-разному. Для добавления такой возможности было предложено множество подходов к добавлению в модель латентной переменной, отвечающей за просодические характеристики речи. К ним относятся как чисто эвристические подходы (глобальные токены стиля, GST), так и обоснованные с вероятностной точки зрения (вариационные автокодировщики). Основная масса таких методов относится к обучению без учителя, не требующему какой-либо просодической разметки данных; существуют и подходы, позволяющие инкорпорировать в архитектуру и процесс обучения как полную, так и частичную разметку.

Основной корпус работ по теме написан исследователями из зарубежных вузов и технологических компаний; отечественной научной литературе вопрос проработан слабо; в то же время, судя по заявлениям российских технологических компаний, подобные модели уже используются в промышленных системах синтеза речи.

Эксперименты по воспроизведению опубликованных работ показали, что разнообразность речи, синтезированной описанными моделями, сильно зависит от используемого речевого корпуса. Также сложность представляет обучение переменной для стиля произнесения, независимой от переменной диктора (в случае многодикторных систем). Эти результаты совпадают с наблюдениями в опубликованной литературе.