

РАЗРАБОТКА ТЕКСТОНЕЗАВИСИМОЙ СИСТЕМЫ ИДЕНТИФИКАЦИИ ПО ГОЛОСУ С ИСПОЛЬЗОВАНИЕМ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ И МНОГОУРОВНЕВОГО ПУЛЛИНГА

Шутов В.В.

(Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А.Бонч-Бруевича)

Научный руководитель – к.т.н., Симончик К.К.
(ООО «ДСП ЛАБС»)

В настоящее время особый интерес направлен на биометрические системы идентификации личности по голосу. Речь является одним из самых доступных источников данных для анализа, а её сбор не требует больших технических усилий. Данные системы могут работать либо в текстозависимом режиме, либо текстонезависимом. Разработка текстонезависимой системы представляет собой наукоемкую и актуальную задачу. Системы идентификации все чаще служат одним из компонентов защиты пользовательских данных, поэтому их все чаще можно встретить в банковских системах или колл-центрах.

Целью работы является построение текстонезависимой системы идентификации по голосу и улучшение последних разработок в этой области. Для достижения поставленной задачи необходимо подготовить базы файлов с голосами различных дикторов, извлечь звуковые признаки MFCC и обучить модель для оценивания принадлежности речи одному из дикторов.

В ходе работы были подготовлены базы звуковых файлов (SRE, SWB, VoxCeleb), изучены основные подходы к решению задачи (GMM-UBM i-vector, DNN-UBM i-vector, TDNN), составлены обучающая и тестовая выборки файлов, продуманы принципиальные схемы обработки звуковых файлов, предложены методы улучшения существующих решений (использование LSTM слоев и многоуровневого пуллинга при извлечении модели голоса диктора, агрессивная аугментация данных) и разработаны первые версии систем с использованием библиотеки Kaldi.

В результате проделанной работы была получена система, позволяющая оценить вероятность принадлежности речевых файлов одному диктору. Было проведено сравнение получившихся результатов с текущими решениями и выявлено что предложенные решения позволяют получать более высокие результаты на открытых базах данных. Дальнейшие исследования позволят улучшить полученные результаты путем усовершенствования методов обработки данных на каждом этапе и модификацией архитектуры нейронной сети.