Добрый день, я студентка группы Б17-501 Баранова Дарья. Тема моей работы – Разработка системы двухфакторной аутентификации на основе голосовой биометрии

Целью работы является разработка системы удаленной аутентификации на основе распознавания речи и идентификации по голосу. Таким образом, аутентификация получается двухфакторной – проверяется фактор знания – то есть правильную ли фразу произносит человек в аудиозаписи – и фактор свойства – похож ли его голос на тот, что сохранен в базе.

В настоящее время аутентификация используется повсеместно, например – в банках, call-центрах, личных интеллектуальных устройствах. Идет постоянное противостояние методов защиты и методов взлома систем, поэтому технологии и способы аутентификации все время совершенствуются. Растущий интерес к биометрическим данным и удаленной аутентификации придает работе еще больше актуальности.

Задача заключается в том, чтобы реализовать отдельно два компонента – распознавания речи и идентификации по голосу, а затем интегрировать их в единую систему. Система будет хранить для каждого пользователя его голос и его фразу-пароль. Когда нужно добавить в систему нового пользователя, создается новая учетная запись, затем сохраняется образец голоса и фраза.

Рассмотрим подробнее компонент распознавания речи. Он необходим, чтобы выделять из аудиофайла ключевую фразу, которую произнес пользователь. Распознавание состоит из следующих этапов. Первый этап – получение акустических представлений.

Удобный способ представить акустические данные - MFCС. Они позволяют учитывать человеческие особенности восприятия звука и вычисляются по алгоритму.

Акустическая модель состоит из рекуррентной нейронной сети. Ее архитектура будет рассмотрена позже. Чтобы избежать задачи сегментирования акустического сигнала будем использовать алгоритм СТС (**Connectionist temporal classification - Коннекционистская временная классификация**). Он заключается в том, чтобы нарезать исходные данные на окна в 10 мс и подавать их на вход сети. Сеть будет выдавать матрицу вероятностей встречи определенного символа алфавита в определенный момент времени.

Пример матрицы выхода СТС представлен на слайде. По вертикали представлены символы алфавита и дополнительный пустой символ, по горизонтали – временные интервалы. Круг символизирует вероятность, что в момент времени встречается символ. Чем темнее круг, тем большую вероятность он отображает. Для декодирования CTC используется алгоритм поиска лучшего пути, и затем операция свертки. Таким образом, акустическая модель уже способна распознавать текст, однако она не обладает никакой информацией на уровне слов и не способна корректировать свою работу исходя из знаний о том, какие слова в языке существуют.

Чтобы улучшить результат, используется языковая модель. На основе нечеткого поиска и специальных метрик будем искать максимально похожее слово в словаре. Метрикой для нечеткого поиска может выступать расстояние Левенштейна и мера жаккара.

* Расстояние левенштейна отражает различность строк.
* Мера Жаккара наоборот отражает похожесть слов. Поэтому из таблицы примеров на слайде можно видеть, что для нечеткого поиска нужно минимизировать расстояние левенштейна и максимизировать меру жаккара.

Для определения оптимальной архитектуры нейронной сети акустической модели сравнивались архитектуры, представленные на графике. Третья сеть из сверточных и реккурентных слоев дала наименьшее значение потерь и продолжала обучение в течение всего времени обучения, поэтому признана оптимальной.

Рассмотрим теперь второй компонент – идентификацию диктора по его голосу. Задача будет решаться через построение оптимальных для каждого диктора моделей гауссовской смеси распределений (Gaussian Mixture Model). На рисунке изображен пример такой модели. Каждой точкой выступают MFCC. Затем строится модель, которая пытается наилучшим образом смоделировать распределение данных. Модель может состоять из одного компонента (слева) или нескольких (справа). Модель будет являться GMM. Именно её мы будем сохранять как образец голоса каждого пользователя.

От количества компонент зависит качество работы. -//- В процессе исследований будут строиться модели с разными параметрами.

- это функция, состоящая из нескольких гауссианов. Другими словами, она (Модель Гауссовой смеси) представляет собой смесь многомерных гауссовых распределений вероятностей (многомерных нормальных распределений), которые наилучшим образом моделируют входной набор данных.

GMM в нашем случае она будет описывать распределение данных и представлять оценку плотности. Голубые окружности показывают смоделированное GMM распределение входных данных.

Однако в использовании GMM есть несколько недостатков. Чтобы их исправить, будем строить одну универсальную фоновую модель UBM (Universal Background Model). По сути она является такой же моделью гауссовой смеси (GMM), но обучена на очень большом объеме данных. После этого уже из нее будем строить GMM для каждого пользователя. для представления дикторонезависимого распределения признаков

Получая запрос на аутентификацию от пользователя, который представляется пользователем 2, компонент идентификации по голосу будет прогонять полученный голос через все модели GMM. Будет определена модель, которая выдала наибольшую оценку похожести. Система сравнит, совпадает ли учетная запись этой модели с тем, кем именно представляется пользователь и выдаст ответ.

описанный алгоритм обучения и построения модели был опробован с различными параметрами. Определена оптимальная по качеству и времени работы модель, она выделена зеленым.

Для оценки точности аутентификации существует показатели точности ………. На их основе было измерено качество работы как компонента идентификации по голосу, так и всей системы в целом.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Перейдем к построению единой системы. Чтобы обладать удобным функционалом, система должна обрабатывать следующие запросы………

Подробное описание реализованных запросов, параметров и ответов от системы представлено на слайде.

Разработка велась на Python 3.7 в данных средах разработки. Использовались данные библиотеки для удобства разработки.

В процессе тестирования реализованной системы были получены следующие показатели точности и значения скорости работы. Из-за двухфакторности системы не был ошибочно допущен ни один пользователь, из-за чего показатель точности равен 100%. Ошибки возникали только в случаях ошибочного отказа. Среднее время обработки запроса на аутентификацию составляет 2,5 секунды

Таким образом, поставленная цель работы была достигнута и была разработана система двухфакторной аутентификации на основе голосовой биометрии. В процессе работы решены все поставленные задачи. Достигнуты следующие показатели точности.

Для оценки точности распознавания существует показатели точности **(Word Error Rate) и (Character Error Rate)**. Они измеряются по одной формуле, представленной на слайде но на разном уровне – на уровне слов и на уровне символов. *Пример распознавания фразы акустической моделью, всей системой представлен на слайде.*

Без языковой модели система дает точность распознавания WER = 40%, CER = 34%, с языковой моделью, т.е вся система целиком уже WER = 33%, CER = 29%, что является достаточно хорошим результатом. (По результатам исследований показатель WER для системы распознавания гугл составляет 18% )