РАЗРАБОТКА ПРОГРАММНО-ТЕХНИЧЕСКОГО МЕТОДА ДЛЯ ОСУЩЕСТВЛЕНИЯ ГОЛОСОВОГО УПРАВЛЕНИЯ

Е. С. Ласточкина

nikakun181@gmai.com;
Научный руководитель — И. Э. Хейдоров, кандидат физико-математических наук, доцент

В статье рассматривается разработанный программно-аппаратный модуль распознавания голосовых команд на базе платформы Raspberry Pi 4 и скрытой марковской модели.

Ключевые слова: распознавание речи; скрытая марковская модель; голосовое управление; Python; Raspberry Pi.

ВВЕДЕНИЕ

Распознавание речи – это многоуровневая задача распознавания образов, в которой акустические сигналы анализируются и структурируются в иерархию структурных элементов (например, фонем), слов, фраз и предложений. Каждый уровень иерархии может предусматривать некоторые временные константы, например, возможные последовательности слов или известные виды произношения, которые позволяют уменьшить количество ошибок распознавания на более низком уровне [1, 2].

СТРУКТУРА РАСПОЗНАВАНИЯ РЕЧИ

В данной работе использовалась следующая структура распознавания речи.

На вход подается необработанная речь, записанная с высокой дискретизацией (около 20 кГц). Определяются границы речи (данный этап требует наличия детекторов речи, отличающих речь от шума по энергии звукового сигнала). Речевой сигнал разделяется на сегменты, длительностью 10–20 мс, и далее происходит процесс выделения векторов признаков, используемых в распознавании.

Для анализа состава речевых кадров требуется набор акустических моделей. В работе использовалась скрытая марковская модель, которая является моделью состояний [1]. После выбора акустической модели сопоставляются различные акустические модели к каждому кадру речи и выдает матрицу сопоставления последовательности кадров. Для моделей, основанных на состоянии, матрица состоит из вероятностей того, что данное состояние может сгенерировать данный кадр. Далее строится алгоритм модели подготовка обучающих данных, извлечение признаков,

разбиение слов на фонемы, обучение модели). Происходит классификация результатов в соответствии с тем, что мы ожидаем на выходе после распознавания.

В результате работы система распознавания речи выдает последовательность слов, которая, наиболее вероятно, соответствует входному потоку речи.

РЕАЛИЗАЦИЯ АКУСТИЧЕСКОЙ МОДЕЛИ ДЛЯ ИЗОЛИРОВАННЫХ СЛОВ

Акустическая модель – это функция, принимающая на вход признаки на небольшом участке акустического сигнала и выдающая распределение вероятностей различных фонем на участке [1]. Таким образом можно по звуку восстановить то, что было произнесено.

Была смоделирована марковская цепь для определения оптимального количества скрытых состояний (фонем произнесенных слов "stop" и "start"). На вход подается последовательность [0, 1, 0, 0, 1, 1, 0, 1], где 0 – "stop", а 1 – "start". В итоге модель определила 5 скрытых состояний (рис. 1).

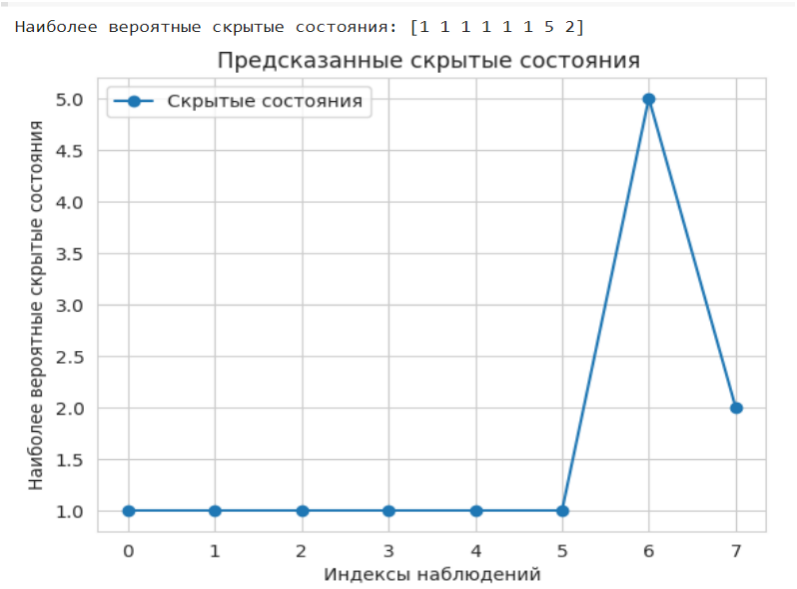

Рис. 1. Результат нахождения оптимального количества скрытых состояний

ВЫБОР АППАРАТНОГО СРЕДСТВА

Для реализации голосового управления был выбран одноплатный компьютер Raspberry Pi 4 Model B 4GB. Данное устройство предоставляет ряд преимуществ для голосового управления. Raspberry Pi 4 обладает более мощным процессором и большим объемом оперативной памяти по сравнению с другими моделями Raspberry Pi. Встроенный графический процессор обеспечивает высокую производительность обработки видео и аудио. Устройство имеет большое количество портов и интерфейсов.

Raspberry Pi 4 гибок в выборе аппаратного обеспечения, что позволяет его эффективно использовать в распознавании речи и интегрировать в более сложные системы.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ РАСПОЗНАВАНИЯ ГОЛОСОВЫХ КОМАНД

Алгоритм распознавания реализован при помощи языка программирования Python и его библиотек numpy, scipy.io, hmmlearn и librosa. Процесс обучения и предсказания осуществляется следующим образом. Задается скрытая марковская модель с определенным количеством компонент, типом ковариации и числом итераций. Загружаются и обрабатываются обучающие данные. Звуковые файлы из разных категорий (папок) читаются, а затем извлекаются акустические признаки, на которых в дальнейшем модель обучается. Далее загружаются и обрабатываются тестовые файлы (каждого файла извлекаются акустические признаки). Для каждого тестового файла сравниваются оценки каждой обученной модели. Индекс модели с наибольшей оценкой определяет предсказанную метку класса для данного тестового файла. В итоге выводится истинная метка класса и предсказанная метка класса для каждого тестового файла (рис. 2).

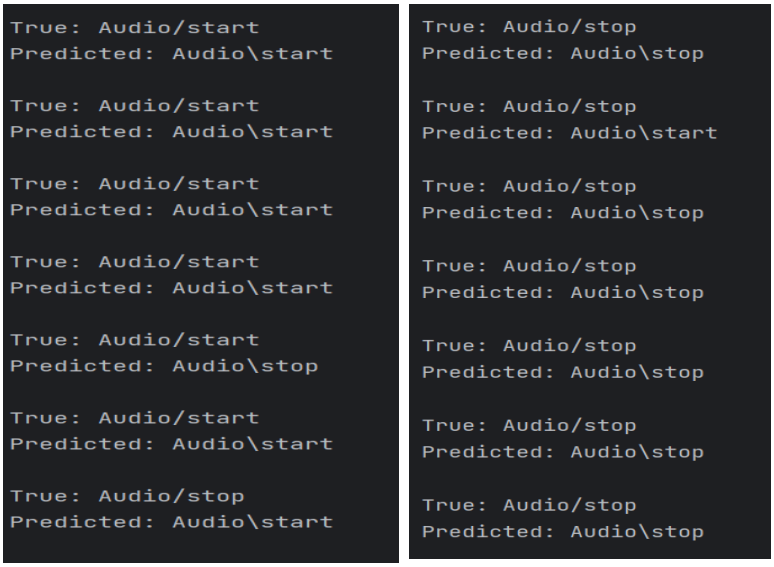

Рис. 2. Результат распознавания голосовых команд "start" и "stop"

ЗАКЛЮЧЕНИЕ

Для реализации распознавания голосовых команд была выбрана акустическая модель. Ею является скрытая марковская модель. Ее преимуществом является простота архитектуры по сравнению с другими популярными моделями распознавания (например, нейронными сетями).

Был разработан подход к реализации голосового управления, анализ количества скрытых состояний при помощи марковской цепи. В дальнейшем выявленное количество скрытых состояний использовалось при реализации скрытой марковской модели.

Разработан и реализован программно-технический метод голосового управления для изолированных слов, реализованный на базе одноплатного компьютера Raspberry Pi 4 с помощью скрытой марковской модели. Разработанный модуль был протестирован на примере аудиозаписей голосовых команд "stop" и "start". Записи различались по длительности, произношению и высоте звука. Результаты показывают, что модуль учитывает данные особенности голосовых команд и успешно их распознает. Таким образом была подтверждена эффективность и надежность разработанного аппаратно-программного модуля на основе метода.

Библиографические ссылки

1. *Рабинер Л.* Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: Обзор / *Л. Рабинер* // Труды института инженеров по электротехнике и радиоэлектронике. – М.: Мир, 1989. Т. 77, № 2. С. 86–120.
2. *Огнев И. В.* Предварительная обработка речевого сигнала для построения базы произношений одиночных слов / *И. В. Огнев, П. А. Парамонов* // Информационные средства и технологии: тр. XX Междунар. науч.-техн. конф. – М. : МЭИ, 2012. С. 53–58.
3. Becchett, C. Speech Recognition. Theory and C++ Implementation / *C. Becchetti, L. P.* Ric otti. Wiley, 1999. 428 c.
4. *Rabiner L.* Fundamentals of speech recognition / *L. Rabiner, B.-H. Juang*. – Prentice Hall, 1993. 507 p.