

УДК 004.934.2

МЕТОДЫ РАСПОЗНАВАНИЯ РЕЧИ В УСЛОВИЯХ ВОЗДЕЙСТВИЯ ШУМОВ

Ф. Ф. оглы МАМЕДОВ

(Представлено: канд. тех. наук, доц. И.Б. БУРАЧЁНОК)

В данной статье рассматриваются эффективные методы распознавания речи в условиях шумов. Установлено, что метод распознавания речи должен быть устойчивым к изменению тембра голоса, громкости и скорости произношения. Определены важнейшие показатели для уникальной идентификации образца по голосу. Рассмотрены наиболее эффективные методы распознавания речи в условиях шумов.

Введение. Задача распознавания речи всегда остается одной из важнейших задач в радиоразведке, криминалистике, антитеррористическом мониторинге, а также для обеспечения безопасности доступа к информационным, финансовым и другим ресурсам. Развитие математического аппарата и совершенствование технических возможностей по его реализации открывает новые возможности при верификации и идентификации человека по голосу. Самым тривиальным методом можно определить сравнение фрагмента речи с фрагментом, текст которого известен заранее и тем самым «распознать» речь. Но имеется ряд условий. Метод распознавания должен быть адаптивным, то есть находится в работоспособности не только при изменении громкости, скорости и тембра, но и быть эффективным в использовании и при технических проблемах, например, при наличии шума или плохого качества звука. Поэтому **целью данной статьи** является рассмотрение известных методов распознавания речи и выявление наиболее эффективного метода распознавания речи в условиях шумов.

В [1] авторами приводится обзор методов распознавания личности по голосу. Указано, что большинство работ по распознаванию диктора сводятся к использованию параметров кепстра (обратного преобразования Фурье от логарифма спектра мощности [2]) и использованию параметров в виде коэффициентов кепстра, который вычисляется по огибающей спектра, полученного через преобразование Фурье, при помощи гребенки соответствующих фильтров, либо по передаточной функции речевого тракта, определенной методом линейного предсказания (Linear Prediction Coding (LPC)). Указывается эффективность при распознавании голоса методов гауссовых смесей (Gaussian mixture models (GMM)) и опорных векторов (Support Vector Machine (SVM)). Приводятся примеры использования скрытых Марковских моделей (Hidden Markov Model (HMM)) и искусственных нейронных сетей (Artificial Neural Networks (ANN)) в задачах распознавания.

Далее подробнее остановимся на некоторых из указанных методов. Рассмотрим их достоинства и недостатки, исследуем эффективность представленных методов в условиях шумов.

Метод на основе скрытых Марковских моделей (HMM). Это один из самых распространенных методов распознавания по голосу, при использовании которого, каждому моделируемому речевому объекту – фразе, слову, слогу, фонеме или аллофону – сопоставляется своя HMM. HMM законченной фразы представляет собой конкатенацию HMM-слов, которые представляются конкатенацией HMM более мелких элементов.

Математическая модель лево-правых HMM. HMM представляет собой итоговый автомат, который изменяет своё состояние в каждый дискретный момент времени t . Переход из состояния S_i в состояние S_j осуществляется случайным образом с вероятностью a_{ij} . В каждый дискретный момент времени модель порождает вектор наблюдений O_t с вероятностью $b_j(O_t)$ [5]. Схема лево-правых HMM без пропусков состояний представлена на рисунке 1.

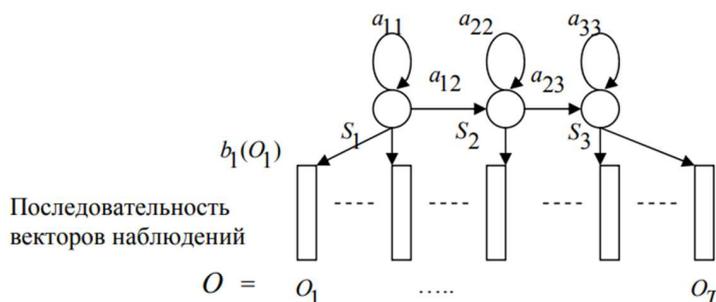


Рисунок 1. – Схема лево-правых HMM без пропусков состояний

Данный алгоритм медленный, однако обладает низкой вычислительной сложностью, что является его преимуществом.

Искусственные нейронные сети (ANN). Структуры и принципы искусственных нейронных сетей основываются на различных взаимосвязанных биологических моделях нервных структур, в том числе в особенности головного мозга. Нейронные сети могут рассматриваться как разновидность самоорганизующихся алгоритмов и представляют собой множество однотипных и параллельно функционирующих элементов или нейронов, связанных между собой и «внешним миром» с помощью специально организованных связей. Самой известной и наиболее распространенной моделью нейронной сети является многослойный перцептрон (МП) [6] (см. рисунок 2).

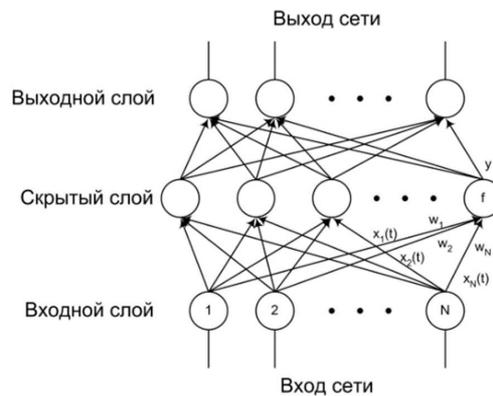


Рисунок 2. – Схема многослойного перцептрона

Элементы многослойного перцептрона разделены на несколько слоев, внутри слоя элементы можно считать линейно упорядоченными и не взаимодействующими между собой. Каждый нейрон сети (кроме нейронов входного слоя – рецепторов) получает входной сигнал от каждого нейрона предыдущего слоя и выходной сигнал нейрона (кроме последнего слоя) поступает на вход нейронов последующего слоя. Таким образом, МП является моделью со связями, обеспечивающими распространение сигнала только вперед (без обратных связей) – от входа к выходу сети. Наиболее известным алгоритмом обучения для МП является алгоритм обратного распространения ошибки, который позволяет осуществить обучение «с учителем» [6].

Нейронная сеть с задерживанием времени (TDNN). TDNN (Time Delay Neural Network) – это сеть, реализующая одну из попыток использования статического МП для распознавания динамической временной последовательности речевых данных путем преобразования временной последовательности в пространственную последовательность соответствующих нейронов.

Как видно из рисунка 3, активность выходного нейрона определяется активностью нейронов из скрытого слоя, взятых в моменты времени $t-1, t, t+1$. Число шагов, на которое МП «заглядывает» вперед и назад во времени, выбирается разработчиком модели. Для обучения сети с такой топологией также может использоваться BP-алгоритм или алгоритм обратного распространения ошибки (Back Propagation Error) [6].

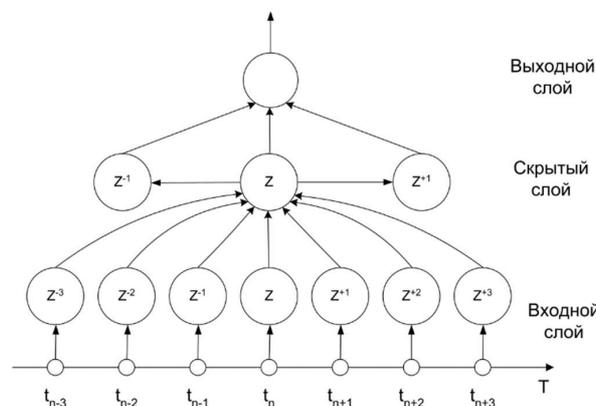


Рисунок 3. – Нейронная сеть с задерживанием времени (TDNN)

Таким образом, при использовании искусственных нейронных сетей имеется возможность параллельной обработки данных, что позволяет при большом количестве межнейронных связей значительно ускорить процесс распознавания и принятия решения. Важным в нейросетевом подходе является возможность обобщения полученных знаний. В результате обучения нейронная сеть способна выявлять сложные зависимости между входными и выходными данными. При обобщении информации такая сеть позволяет получить наиболее верный результат на основании неполных или искажённых данных.

Метод динамического программирования. Суть метода состоит в том, чтобы сравнить две совокупности векторов разной длины, причем на пространстве векторов есть метрика или мера близости. Предположим, что мы сравниваем эталон сам с собой: отложим векторы признаков эталона по оси X и Y . На плоскости XU на пересечении координат, соответствующих векторам i и j , построим вертикальный отрезок, равный расстоянию (степени близости) между этими векторами. Тогда на квадрате со стороной, равной количеству векторов в эталоне (N), возникнет «гористый ландшафт», симметричный относительно диагонали $(0,0) (N, N)$, однако по диагонали будет пролегать абсолютно прямая «долина» с высотой, равной 0 (исходя из того, что расстояние от вектора до самого себя равно 0). В случае, если мы сравниваем два разных эталона, принадлежащих одному слову, то картина исказится, однако, если используемые признаки адекватно отражают процесс восприятия, можно надеяться, что некоторая «долина» по-прежнему будет пролегать по ломаной, близкой к диагонали, теперь уже прямоугольника [7]. Схема метода динамического программирования представлена на рисунке 4.

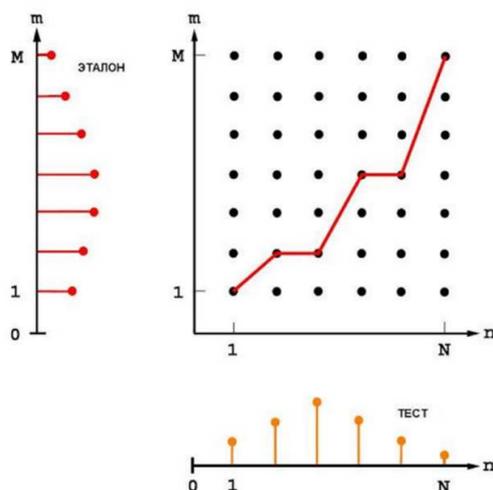


Рисунок 4. – Схема метода динамического программирования

При использовании представленного метода для придания системе дикторонезависимых качеств, для каждого слова записывают несколько эталонов от разных дикторов, а в процессе обучения если эталон диктора не распознан, то его добавляют в базу эталонов. Однако данный алгоритм не решает задачу определения начала/конца слова и позволяет в дикторозависимом варианте распознавать 100-300 слов с вероятностью 90-98% [7].

Вывод. В результате анализа эффективности представленных методов для распознавания диктора по голосу в условиях шумов выявлено, что:

- при использовании искусственных нейронных сетей имеется возможность значительно ускорить обработку данных для получения наиболее верного результата по распознаванию
- при использовании метода на основе скрытых Марковских моделей на примере рассмотренного алгоритма лево-правых НММ без пропусков состояний установлено, что данный алгоритм требует больших временных затрат. Однако он обладает низкой вычислительной сложностью, что иногда является ключевым при решении задач распознавания;
- при использовании метода динамического программирования основная проблема – различный темп и нелинейность темпа произнесения фрагмента речи. Это простотой метод, обладающий высоким быстродействием и отсутствием необходимости собирать речевую базу данных.

Таким образом, можно сделать вывод, что задача выбора эффективного метода распознавания диктора решается определением компромисса между производительностью, требованиям к ресурсам и качеством работы системы.

ЛИТЕРАТУРА

1. Сорокин, В.Н. Распознавание личности по голосу: аналитический обзор / В.Н. Сорокин, В.В. Вьюгин, А.А. Тананыкин // Информационные процессы. – 2012. – Т. 12, № 1. – С. 1–30.
2. Гоноровский И. С. Радиотехнические цепи и сигналы: Учебник для вузов. – 4-е изд., перераб. и доп. – М.: Радио и связь, 1986. – 512 с.
3. Жилияков, Е. Г. Сегментация речевых сигналов на основе субполосного анализа / Е. Г. Жилияков, А. А. Фирсова // Вестник Нац. техн. ун-та «ХПИ»: сб. науч. тр. Темат. вып. : Информатика и моделирование. – Харьков : НТУ «ХПИ». – 2013. – № 39. – С. 73-81.
4. Джалолов, У.Х. Регуляризация задачи идентификации объекта в условиях зашумленности полезного сигнала / У.Х. Джалолов, Р.М. Бандишоева, У.А. Турсунбадалов // Вестник Таджикского технического университета. – 2016. – № 1 (33). – С. 20–26.
5. Ахмад, Х. М. Введение в цифровую обработку речевых сигналов : учеб, пособие / Х. М. Ахмад, В. Ф. Жирков ; Владим. гос. ун-т. – Владимир : Изд-во Владим. гос. ун-та, 2007. – 192 с.
6. Модели, методы, алгоритмы и архитектуры систем распознавания речи.; под редакцией В.В. Рязанова. – М.: Российская академия наук, Вычислительный центр им. А.А. Дородницына, 2006. – 138 с.
7. Тампель, И.Б., Карпов А.А. Автоматическое распознавание речи: Учебное пособие. / И.Б. Тампель, А.А. Карпов. – Санкт-Петербург: Университет ИТМО, 2017. – 152 с.