

Список использованных источников:

1. Chandra, Dr.E. Keyword Spotting: An Audio Mining Technique in Speech Processing / Dr.E. Chandra, K.A. Senthildevi // IOSR Journal of VLSI and Signal Processing. – 2015. – №5. – с. 22-27.
2. Barakat, M. S. Keyword spotting based on the analysis of template matching distances / M. S. Barakat, C. H. Ritz, D. A. Stirling // Signal Processing and Communication Systems. – 2011. – с. 1-6.
3. John Sahaya Rani Alex. Spoken Utterance Detection Using Dynamic Time Warping Method Along With a Hashing Technique / John Sahaya Rani Alex, Nithya Venkatesan // International Journal of Engineering and Technology. – 2014. – №2 (6). – с. 1100-1108.
4. Zheng Fang Comparison of Different Implementations of MFCC / Zheng Fang, Zhang Guoliang, Song Zhanjiang // J. Comput. Sci. A Technol. – 2011. – №6 (16). – с. 582-589.
5. Eamonn, J.K. Derivative Dynamic Time Warping / J.K. Eamonn, J.P Michael // SIAM International Conference on Data Mining. – 2001. – с. 11.

СИСТЕМА ИДЕНТИФИКАЦИИ ДИКТОРА ПО ОБРАЗЦУ ГОЛОСА НА БАЗЕ РЕЧЕВЫХ ЗАПИСЕЙ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Вашкевич Г. С.

Вашкевич М. И. – канд. техн. наук, доцент

В данной работе описывается общий подход к построению систем идентификации диктора, проводится обзор современных методов и подходов к решению данной задачи, а также представляется пример реализации системы идентификации диктора на основе векторного квантования.

В условиях интенсивного развития систем распознавания речи все чаще перед разработчиками стоит задача не только распознать речь, но и установить личность диктора. Такие системы нашли широкое применение в криминалистике, системах контроля доступа и верификации, а также в обработке речевых баз данных.

На данный момент основными методами идентификации дикторов являются методы, основанные на следующих моделях: модели гауссовых смесей (англ. Gaussian Mixture Models, GMM), модель, основанная на методе опорных векторов (англ. support vector machine, SVM), и модель, основанная на векторном квантовании [1].

В методе, использующем гауссовы смеси, модель голоса диктора представляет собой аппроксимацию распределения используемых речевых признаков смесью гауссовых распределений. Недостатком данного метода является его вычислительная сложность, и сложность реализации [2].

Метод опорных векторов используется для разделения сложных областей в пространстве признаков речевого сигнала на два класса оптимальной (с точки зрения алгоритма) гиперплоскостью. В этом методе отыскиваются образцы, находящиеся на границах между двумя классами и через них строятся две параллельные гиперплоскости, с максимальным расстоянием между ними. После этого на равном удалении от каждой из найденных гиперплоскостей строится разделяющая гиперплоскость. Таким образом все обучающие вектора модели диктора находятся по одну сторону от разделяющей гиперплоскости. Если на вход системы подать тестовый вектор, то система будет сравнивать его положение в пространстве с положением гиперплоскости. Если тестовый вектор находится по одну сторону от гиперплоскости вместе с обучающими векторами модели диктора – то система принимает решение о том, что данный тестовый характеристический вектор принадлежит диктору.

Модель, основанная на векторном квантовании, строит модель диктора на основе кодовой книги, полученной путем векторного квантования обучающей выборки характеристических векторов диктора. Векторное квантование в общем случае осуществляется при помощи алгоритма кластеризации K-средних. Принятие решения о принадлежности тестового характеристического вектора диктору осуществляется при превышении порогового значения меры близости между входным вектором и, хотя бы одним вектором из кодовой книги.

В общем случае системы идентификации дикторов имеют структуру, представленную на рисунке 1:

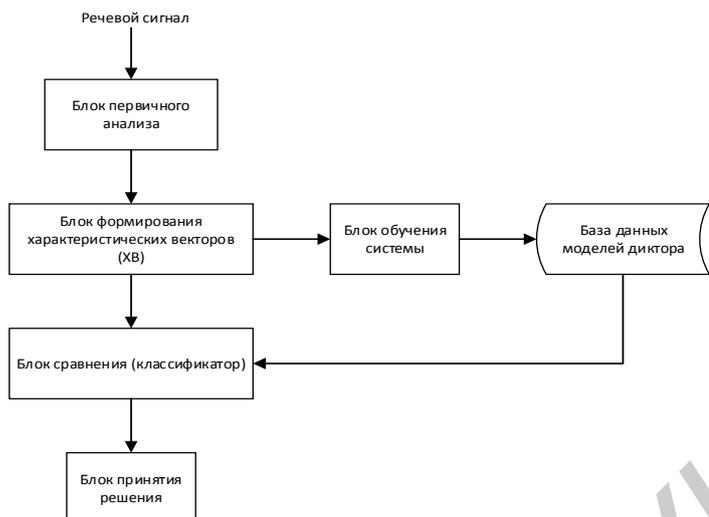


Рис. 1. Структурная схема системы идентификации диктора

Блок первичного анализа предназначен для анализа входного сигнала (разбиения речевого сигнала на фреймы) и выделения участков записи, на которых присутствует речь. Блок формирования характеристических векторов из полученных фреймов формирует мел-частотные кепстральные коэффициенты (MFCC), которые используются для обучения системы в соответствующем блоке. Блок сравнения сравнивает модели дикторов, которые хранятся в базе системы с характеристическими векторами входного сигнала. Результат сравнения передается в блок принятия решения, который сравнивает результаты работы классификатора с пороговым значением и формирует результат идентификации.

Для реализации системы идентификации диктора по образцу голоса был выбран метод, основанный на алгоритме векторного квантования.

Для построения модели диктора используется модифицированный алгоритм К-средних, который состоит из следующих шагов:

1. Ввод обучающих векторов

$$S = x_i \in R^d \quad i = 1, 2, \dots, n .$$

2. Определяется начальное состояние вычисляемой кодовой книги

$$C = c_i \in R^d \quad i = 1, 2, \dots, K .$$

3. Установить $D_0=0$ и $k = 0$.

4. Классифицировать обучающих векторов в кластеры в соответствии с правилом

$$x_i \in S_q \text{ if } x_i - c_{q-p} < x_i - c_{j-p} \text{ для } j \neq q .$$

5. Обновить кластеры $c_j, j=1, 2, \dots, K$ как

$$c_o = \frac{1}{S_j} \sum_{x_i \in S_j} x_i .$$

6. Установить $k = k+1$ и вычислить

$$D_k = \sum_{j=1}^K \sum_{x_i \in S_j} x_i - c_{j-p} .$$

7. Если $\frac{D_{k-1} - D_k}{D_k} > \varepsilon$, повторить шаги 4 – 6.

8. Вернуть текущую кодовую книгу

$$y C = c_i \in R^d \quad i = 1, 2, \dots, K .$$

Модификация базового алгоритма К-средних выполняется с целью уменьшения влияния выбранных начальных положений центроидов кластеров. Выбор начального состояния центроидов следует проводить по следующему алгоритму:

1. Выбрать первый центроид случайным образом (среди всех точек).
2. Для каждой точки найти значение квадрата расстояния до ближайшего центроида (из тех, которые уже выбраны) dx^2 .
3. Выбрать из этих точек следующий центроид так, чтобы вероятность выбора точки была пропорциональна вычисленному для неё квадрату расстояния. Это можно сделать следующим образом. На шаге 2 нужно параллельно с расчётом dx^2 подсчитывать сумму $Sum(dx^2)$. После накопления суммы найти значение $Rnd = random(0.0, 1.0) * Sum$. Rnd случайным образом укажет на число из интервала $[0; Sum)$, и нам остаётся только определить, какой точке это соответствует. Для этого нужно снова начать подсчитывать сумму $Sum(dx^2)$ до тех пор, пока сумма не превысит Rnd . Как только это случится, суммирование останавливается, и мы можем взять текущую точку в качестве центроида. При выборе каждого следующего центроида специально следить за тем, чтобы он не совпал с одной из уже выбранных в качестве центроидов точек, не нужно, так как вероятность

повторного выбора некоторой точки равна 0.

4. Повторять шаги 2 и 3 до тех пор, пока не будут найдены все необходимые центроиды.

Таким образом полученная кодовая книга представляет собой упрощенную модель диктора, что позволяет сократить количество используемой памяти в сравнении с моделью, состоящей из всех характеристических векторов обучающей выборки, а также позволит повысить скорость вычислений.

Процесс работы системы, реализованной по такому алгоритму сводится к полному перебору всех векторов кодовых книг и их попарному сравнению с входным характеристическим вектором. Если в ходе сравнения обнаружится, что мера близости между векторами превышает определенное пороговое значение – то системой принимается решение о принадлежности входного речевого сигнала к диктору, чья модель участвовала в сравнении.

Важной особенностью работы такой системы является возможность добавления новых моделей дикторов в уже функционирующую систему, не переобучая её. Для того, чтобы система понимала нового диктора достаточно в базу данных, хранящую модели дикторов загрузить новую которую книгу.

Рассмотренный в данной статье метод идентификации дикторов обладает целым рядом преимуществ:

- небольшое потребление памяти для хранения моделей дикторов;
- высокая скорость работы;
- простота добавления новых моделей дикторов в уже функционирующую систему.

Список использованных источников:

1. Первушин, Е. А. Обзор основных методов распознавания дикторов / Е. А. Первушин // Математические структуры и моделирование. – 2011. – №24. – С. 41–54.
2. Садыхов Р. Х., Ракуш В. В. Модели гауссовых смесей для верификации диктора по произвольной речи // Доклады БГУИР. – 2003. – №4. – С. 95–103.