УДК 004.93.1

МЕТОДЫ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО ГОЛОСУ, ИХ ПРЕИМУЩЕСТВА И НЕДОСТАТКИ

Минаева И.А.[1], Белов Ю.С.[1]

[1]Калужский филиал ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», г. Калуга (248000, г. Калуга, ул. Баженова, д. 2.), e-mail: liramiks@yandex.ru

К настоящему моменту существует большое количество программного обеспечения, которое позволяет идентифицировать пользователя по голосу. В зависимости от области применения, особенностей распознавания используются различные методы идентификации. Несмотря на отличие методов идентификации, можно выделить общий алгоритм распознавания пользователя по голосу, который включает в себя: извлечение особенностей речи и построение шаблона говорящего, который основывается на признаках, полученных при извлечении речевого сигнала. Каждый метод подходит под определенные условия, в которых происходит идентификация. В данной статье рассматриваются примеры основных методов идентификации, особенности каждого метода, а также математическое представление методов, определяются условия, в которых стоит применить данный метод, приводятся преимущества и недостатки пяти методов: алгоритм динамической трансформации временной шкалы (составляется путь наименьшей стоимости, который должен удовлетворять условиям непрерывности, монотонности, а также граничным условиям), скрытые Марковские модели (преобразование речи в мел-частотные кепстральные коэффициенты), векторное квантование (на основе словаря p-мерных векторов), метод опорных векторов (применим для различия объектов двух классов), модель гауссовых смесей (любой говорящий относится к определенной модели, основанной на гауссовых смесях).

Ключевые слова: идентификация личности по голосу, динамическая трансформация временной шкалы, скрытые Марковские модели, мел-частотные кепстральные коэффициенты, векторное квантование, метод опорных векторов, модель гауссовых смесей

METHODS OF IDENTIFICATION OF PERSONALITY BY VOTE, THEIR ADVANTAGES AND DISADVANTAGES

Minaeva I.A.[1], Belov Yu.S.[1]

[1]Moscow State Technical University N.A. Bauman (National Research University), Kaluga Branch, Kaluga, e-mail:
liramiks@yandex.ru

To date, there are a large number of software that allows you to identify the user by voice. Depending on the application and recognition features, various identification methods are used. Despite the difference in identification methods, it is possible to single out a general algorithm for recognizing a user by voice, which includes: extracting speech features and building a speaker pattern that is based on features obtained when extracting a speech signal. Each method is suitable for certain conditions in which the identification takes place. This article discusses examples of basic identification methods, features of each method, as well as a mathematical representation of the methods, identifies the conditions in which this method should be applied, the advantages and disadvantages of the five methods: the algorithm for dynamic transformation of the time scale conditions of continuity, monotony, as well as boundary conditions), hidden Markov models (conversion of speech into mel-frequency cepstral coefficients s), vector quantization (on the basis of the dictionary of p-dimensional vectors), the method of support vectors (applicable to distinguish objects of two classes), the model of Gaussian mixtures (any speaker refers to a specific model based on Gaussian mixtures).

Keywords: voice identification, dynamic transformation of the time scale, hidden Markov models, mel frequency cepstral coefficients, vector quantization, support vector machine, Gaussian mixture model

Введение. В данной статье рассматриваются существующие способы автоматической идентификации личности, основываясь на голосе говорящего человека. Невзирая на отличие, каждому из методов идентификации свойственны общие этапы, среди которых можно выделить:

1. Получение речевых характеристик звукового сигнала.

2. Построение шаблона говорящего, основываясь на полученных признаках.

Идентификация говорящего в какой-либо системе по голосовому вводу заключается в поиске самой подходящей модели, основываясь на определенных критериях.[1]

Алгоритм динамической трансформации временной шкалы (DTW).данный алгоритм дает возможность найти схожесть между распознаваемой звуковой последовательностью и сравниваемым эталонным образцом, взяв во внимание изменения, происходящие во времени, путем вычисления последовательности, которая будет наиболее оптимальна для изменения во времени для двух последовательностей векторов, содержащих речевые признаки: G – последовательность извлеченных речевых признаков из обучающей выборки и P – из тестовой выборки[2,3]:

$$G=\{g_1,g_2,\ldots,g_n\},\quad P=\{p_1,p_2,\ldots,p_m\} \qquad (1)$$

Основываясь на (1) составляется путь наименьшей стоимости.

Для построения матрицы $\Omega^{n\cdot m}$ необходимо найти расстояние между i-ым и j-ым элементами из (1). Это расстояние и будет являться элементами матрицы $\Omega^{n\cdot m}$.По матрице $\Omega^{n\cdot m}$ строится путь W, показывающий соответствие между исходными последовательностями. k-ый элемент W

определяется как $w_k = (i, j)$, где К должен соответствовать ограничению:

$$min(m,n)\leq K<m+n-1 \qquad (2)$$

Путь W должен строиться на основе следующих условий:

- Непрерывность. При проходе по последовательности i и j изменяются только на единицу, то есть проход совершается пошагово. Таким образом, в шаге пути принимают участие соседние элементы.
- Монотонность. Если $w_k = (a,b)$ и $w_{k-1} = (p,g)$, тогда $a-p\geq 0,\ b-g\geq 0$. Это нужно для того, чтобы точки пути размеренно перемещались во времени.[8]
- Диагональ матрицы должна содержать начало и конец пути, которые располагаются в противоположных углах, что представляет собой граничные условия.

Для выбора пути с наименьшей стоимостью (выравнивающий путь) необходимо воспользоваться формулой:

$$DTW(G,P) = min\left\{\frac{1}{K}\sqrt{\sum_{k=1}^{K} d(w_k)}\right. \qquad (3)$$

Знаменатель К нужен для учета разной длины пути.

Таким образом, искомый W для взятых последовательностей представляет собой путь, при котором образовывается минимальная стоимость *DTW(G,P)*.

Для идентификации говорящего определяется минимальная стоимость для всех шаблонов из базы данных. Значение, которому соответствует путь с минимальной длиной, определяет n - индивидуальный номер диктора, чей голос максимально приближен к образцу исходной речи. К преимуществам данного метода можно отнести легкую реализацию. К недостаткам можно отнести невозможность применения алгоритма для текстонезависимой идентификации, что делает его почти неиспользуемым в современных системах идентификации.

Скрытые Марковские модели (СММ). СММ – модель, которая описывает стохастический процесс, который разбивается на несколько этапов. Первый этап включает в себя создание цепи Маркова. Второй этап включает в себя создание временной последовательности для каждой точки цепи Маркова. Эта последовательность является выходной. Наблюдателю процесса доступна временная последовательность. В процессе генерации выделяется последовательность состояний, которая недоступна наблюдателю процесса («скрыта») [8].

Модель Маркова характеризуется следующими элементами:

- Множество, состоящее из скрытых состояний;
- Множество, состоящее из наблюдений;
- распределение состояний, определяющая вероятность начать работу в определенном состоянии i;
- матрица вероятностей переходов между скрытыми состояниями
- матрица, содержащая вероятности

Для идентификации диктора по голосу предпринимаются следующие этапы: речевой сигнал преобразуется в мел-частотные кепстральные коэффициенты. После чего применяется алгоритм векторного квантования к полученным коэффициентам. В результате применения алгоритма квантования вычисляется последовательность наблюдений $O=\{o_1,o_2\ldots,o_t\}$. С помощью полученной на предыдущем этапе последовательности наблюдений, параметров моделей пользователей λi рассчитывается вероятность $P(O|\lambda i)$, которая определяет процент совпадения последовательности O и модели пользователя. К преимуществам данного метода можно отнести улучшенное качество распознавания; возможность достаточно скоро

восстанавливать порядок состояний модели с помощью информации о длительности каждого состояния. К недостаткам метода можно отнести большое количество вычислений и большое количество памяти. Обязательная оценка большого количества новых параметров – состояний также можно отнести к недостаткам метода[4].

Векторное квантование. Векторное квантование (VQ) упрощает распознавание за счет сжатия сигнала речи. Первый этап (обучение) предполагает создание словаря, содержащего p-мерные вектора (эталонные слова). На втором этапе (классификация) для всех векторов из выборки тестовых векторов s_i определяются k соседних кодовых векторов. Далее тестовый вектор замещается индексом максимально приближенным к кодовому слову.

Вероятность принадлежности вектора s диктору D можно рассчитать по формуле:

$$P(D_j|s_i) = \frac{k_{ij}}{k} \qquad (4)$$

Классификация последовательность тестовых векторов находится по формуле:

$$S = argmax \prod_{i=1}^{L} P(D_j|s_i); 1 \le j \le N \qquad (5)$$

Таким образом, векторное квантование можно разбить на следующие стадии:

1. Обучение. Включает инициализацию, поиск ближайшего кодового слова, обновление кодовой книги, итерация

2. Классификация. На вход подается неизвестный вектор. После соответствующих преобразований получают индекс кодового слова, ближайшего к входному вектору [6].

Среди преимуществ данного метода можно выделить следующие: метод подходит для задач текстонезависимой идентификации диктора; метод прост в программном исполнении. Среди недостатков можно выделить следующее: метод дает не всегда высокую точность идентификации.

Метод опорных векторов (SVM). Данный алгоритм применятся для решения одной задачи: различать объекты двух классов, причем, делая это гораздо быстрее, чем нейронные сетеи. Метод использует функцию разделения:

f(x)=w*x+b

Пусть X - последовательность точек пространства признаков, Y- значения, которые описывают два класса.

Говоря о данных, можно выделить линейно-разделимые и линейно-неразделимые. В первом случае условия можно записать так:

$$\begin{cases} w * x_i + b \ge 1, y_i = 1 \\ w * x_i + b \le -1, \ y_i = -1 \end{cases} \qquad (6)$$

Для надежного разделения классов нужно, чтобы расстояние между разделяющими гиперплоскостями было наибольшим, что можно реализовать при помощи метода множителей Лагранжа [6].

Во-втором случае вводится функция ядра. Чтобы работать с линейно неразделимым множеством, для того, чтобы отразить текущее пространство в пространство с большей размерностью нужно задать функцию ядра:

$$K(x_i, x_j) = \varphi(x_i) * \varphi(x_j) \qquad (7)$$

в полученном пространстве данные можно линейно разделить.

После вычисления функции f(x), принадлежность вектора x’ соответствующему классу определяется знаком выраженияf(x’)

Задачи мультиклассового распознавания строятся на стратегии «один против всех». Создается ряд классификаторов g. Каждый классификатор может отличать один определенный класс от всех остальных. При идентификации объект заносится в тот класс, чей классификатор показал максимальное значение f(x). К преимуществам данного метода можно отнести то, что метод достигает высокую точность классификации; при изменении функции ядра метод может использовать разные методы классификации. При мультиклассовой идентификации обучение протекает медленно, что можно отнести к главному недостатку метода.

Модель гауссовых смесей GMM. GMM- это взвешенная сумма *M* компонент:

$$P(\bar{x} | \lambda) = \sum_{i=1}^{M} p_i b_i(\bar{x}) \qquad (8)$$

где$\bar{x}$ – D-мерный вектор случайных величин, p_i– веса компонентов модели, b_i– функции плотности распределения. Для p_i (весов) должно выполняться условие:

$$\sum_{i=1}^{M} p_i = 1 \qquad (9)$$

Полностью GMM можно определить векторами математического ожидания, ковариационными матрицами и весами смесей для каждого компонента модели:

$$\lambda = \left\{ p_i \mu_i \sum i \right\}, i = 1, \dots M \quad (10)$$

Используя метод GMM, любой диктор - λ является моделью гауссовых смесей.[8]

При реализации модели говорящего нужно оценить её параметры, таким образом, чтобы они были максимально близки к распределению векторов обучающей речи.

Для быстрого достижения результатов используется метод оценки максимального правдоподобия. Главная задача этого метода – это найти такие параметры модели, которые максимально правдоподобны этой модели, при заданных обучающих данных [7]. К преимуществам данного метода можно отнести возможность

моделирования большого числа индивидуальных акустических признаков речи диктора. Среди недостатков выделяют возникновение проблемы неустойчивости выборочных оценок плотности и самого классификатора при обращении ковариационных матриц, которые могут быть вырожденными.

Заключение. Идентификация личности по голосу удобна в применении. Главное при построении системы идентификации по голосу – выбор параметров, которые являются индивидуальными для каждого говорящего. В данной статье были рассмотрены методы идентификации, применяемые при различных условиях распознавания.

Список литературы

1. Васильев Р.А. Исследование особенностей идентификации дикторов по голосу. Известия ТулГУ. Технические науки. 2013. Вып. 3. С. 246-248

2. Гришунов С.С., Молчанов А.Н., Бурмистров А.В. К вопросу об эффективности систем верификации пользователей по голосу. Электронный журнал: наука, техника и образование. 2017. № 1 (10). С. 16-20.

3. Гришунов С.С., Бурмистров А.В., Молчанов А.Н. Математические методы классификации дикторов. Вопросы радиоэлектроники. 2016. № 10. С. 13-17.

4. Jaafer S. A., Mohamed I. O., Elhafiz M. M. Text-Independent Speaker Identification Using Hidden Markov Model. World of Computer Science and Information Technology Journal (WCSIT), 2012, vol. 2, pp. 203-208

5. Сухаревская Е.В. Исследование систем аутентификации // Международный студенческий научный вестник. 2018. № 1.; URL: http://eduherald.ru/ru/article/view?id=18090 (дата обращения: 01.11.2018).

6. Нифонтов С.В, Белов Ю.С. Применение скрытых марковских моделей в текстонезависимых системах идентификации пользователей по голосу // Наука, техника и образование. 2016 №2. С. 116-124

7. Ковтун В.В., Федоров Е.Е. Разработка компьютерной системы распознавания речи // Современные научные исследования и инновации. 2015. № 1. Ч. 1. URL http://web.snauka.ru/issues/2015/01/45410 (дата обращения: 01.11.2018).

8. Запрягаев С. А., Коновалов А. Ю. Распознавание речевых сигналов // Вестник ВГУ, № 2, 2009, С. 39–48.