

Аль Мандили Мухамад Р.А., Лебедев В.В., Чернышев О.Л. Описание алгоритма голосовой аутентификации пользователей. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей XIX Междунар. научно-техн. конф. – Пенза: ПДЗ, 2019. – С. 023-027.

УДК 004.056.5

ОПИСАНИЕ АЛГОРИТМА ГОЛОСОВОЙ АУТЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ

Р.А. Аль Мандили Мухамад, В.В. Лебедев, О.Л. Чернышев

DESCRIPTION OF ALGORITHM FOR VOICE AUTHENTICATION OF USERS

R.A. Al Mandili Muhamad, V.V. Lebedev, O.L. Chernyshev

Аннотация. Статья посвящена описанию процедуры голосовой аутентификации пользователей, связанной с определением вектора характеристик голоса посредством вычисления кепстральных коэффициентов и его использования для решения задач фильтрации по отпечаткам голоса на основе метода Гауссовских смешанных моделей.

Ключевые слова: отпечаток голоса, кепстральные коэффициенты, аутентификации, фильтр по отпечаткам голоса.

Abstract. The article describes the procedure for voice authentication of users associated with determining the vector of voice characteristics by calculating cepstral coefficients and using it to solve filtering tasks by voice prints based on the Gaussian mixed model method.

Keywords: voice imprint, cepstral coefficients, authentication, voice imprint filter.

Подходы к разработке средств анализа звукового потока с целью определения характеристик голоса человека для последующей его аутентификации уже были детально рассмотрены в публикации авторов [7]. Далее приведем описание алгоритма голосовой аутентификации пользователей.

Для распознавания голоса говорящего рассматривается самый распространенный метод Гауссовских смешанных моделей [4, с. 3-7]. Данный метод выбран F. Bimbot [3, с.7-10] и D.A. Reynolds, T.F. Quatieri, R.B. Dunn [4, с.2-7] как основной для решения задач с вероятностными решениями, в частности, для работы с недетерминированной обработкой звука.

A. Al Marashli, O. Al Dakkak используют многомерную Гауссовскую вероятностную функцию плотности (ВФП) для отображения характеристик процесса обработки звука. Для каждого звукового класса используется своя ВФП, которую можно вычислить по формуле

$$b_i(x) = \frac{1}{(2\pi)^{\frac{R}{2}} \cdot |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^T(\Sigma_i)^{-1}(x-\mu_i)},$$

где Σ_i – ковариационная матрица ВФП; μ_i – вектор средних значений для ВФП; x – вектор кепстральных коэффициентов отдельно взятого говорящего.

Каждый звуковой класс изображается конкретной ВФП функцией. Звучащий речевой поток каждого говорящего отображается группой ВФП в зависимости от того, во сколько классов звука попадает данный голос. Такие группы обозначаются символами $\lambda_1, \lambda_2, \lambda_3, \dots$ и будут условно называться «модели говорящих». Модели говорящих нужно построить и оптимизировать путем тренировки этих моделей:

во-первых, задать произвольные значения для ВФП и весовых коэффициентов p_i в зависимости от векторов характеристик;

во-вторых, использовать алгоритм Expectation Maximization [1, с.3], с помощью которого можно оптимизировать Гауссовские смешанные модели путем увеличения значений условной вероятности $P(x_m|\lambda_s)$ – вероятности попадания вектора характеристик x_m в звуковую модель λ_s , связанной с говорящим под индексом s .

«Expectation-maximization» – EM-алгоритм, используемый в математической статистике для нахождения оценок максимального правдоподобия параметров вероятностных моделей в случае, когда модель зависит от некоторых скрытых переменных. Каждая итерация алгоритма состоит из двух шагов. На E-шаге (expectation) вычисляется ожидаемое значение функции правдоподобия, при этом скрытые переменные рассматриваются как наблюдаемые. На M-шаге (maximization) вычисляется оценка максимального правдоподобия, таким образом увеличивается ожидаемое правдоподобие, вычисляемое на E-шаге. Затем это значение используется для E-шага на следующей итерации. Алгоритм выполняется, пока актуально следующее неравенство [6, с.3]:

$$p(X_{train}|\lambda^{k+1}) \geq p(X_{train}|\lambda^k),$$

где X_{train} – векторы кепстральных коэффициентов для тренировки моделей, λ^{k+1} – модель, полученная после $k + 1$ циклов тренировки.

При условной вероятности

$$p(i_n = i | x_n, \lambda^k) = \frac{p_i^k b_i^k x_n}{\sum_{i=1}^I p_i^k b_i^k x_n}$$

описать алгоритм Expectation Maximization можно следующим комплексом уравнений [6, с.3], [1, с.4]:

$$p_i^{k+1} = \frac{1}{M_{train}} \sum_{n=1}^{M_{train}} p(i_n = i | x_n, \lambda^k)$$

$$\mu_i^{k+1} = \frac{\sum_{n=1}^{M_{train}} p(i_n = i | x_n, \lambda^k) \cdot x_n}{\sum_{n=1}^{M_{train}} p(i_n = i | x_n, \lambda^k)}$$

$$\Sigma_i^{k+1} = \frac{\sum_{n=1}^{M_{train}} p(i_n = i | x_n, \lambda^k) \cdot x_n \cdot x_n^T}{\sum_{n=1}^{M_{train}} p(i_n = i | x_n, \lambda^k)} - \mu_i^{k+1} \cdot (\mu_i^{k+1})^T,$$

где M_{train} – количество моделей; μ_i^{k+1} – среднее значение i -го Гауссовского распределения; Σ_i^{k+1} – ковариационная матрица i -го Гауссовского распределения; x_n – n -й компонент вектора кепстральных коэффициентов.

По окончании процесса тренировки могут быть получены модели говорящих, которые можно использовать для распознавания голоса определенного человека. Вероятность попадания в группу нового голоса с вектором характеристик (X) определена формулой [1, с.2]

$$p(x|\lambda) = \sum_{i=1}^l p_i b_i(x),$$

где p_i – весовой коэффициент; $b_i(x)$ – ВФП i -го класса.

Для распознавания говорящего требуется вычисление значения вероятности сходства конкретного голоса с моделями всех говорящих, которые могут быть построены методом Гауссовских смешанных моделей по характеристикам голосов, затем необходимо выбрать модели с самой высокой вероятностью сходства по формуле [1, с. 3]:

$$\hat{S} = \text{Ind} \left\{ \max_{1 \leq s \leq S} [P(\lambda_s | x_n)] \right\},$$

где \hat{S} – результат выбора (говорящий, который идентифицирован по входящему вектору кепстральных коэффициентов).

Чтобы вывести значения вышестоящей формулы, может быть применено правило Байеса, в соответствии с которым необходимо найти значения $p(x_n | \lambda_j)$. Если вектор характеристик будет состоять из числа M независимых друг от друга характеристик, то по следующей формуле можно найти модель голоса говорящего с максимальной вероятностью:

$$p(\{x_1, \dots, x_M\} | \lambda_j) = \prod_{m=1}^M p(x_m | \lambda_j).$$

Подтверждение того, что отрывок речи принадлежит гипотетическому говорящему (S), может быть найдено путем вычисления по формулам детектора максимального сходства:

$$\frac{p(\{x_1, \dots, x_m\} | H_0)}{p(\{x_1, \dots, x_m\} | H_1)} = \Lambda \begin{cases} \leq \theta \rightarrow \text{accept } H_0 \\ > \theta \rightarrow \text{reject } H_0 \end{cases}$$

$\left. \begin{array}{l} H_0 = (x_m)_{1 \leq m \leq M} \text{ belong to } S \\ H_1 = (x_m)_{1 \leq m \leq M} \text{ doesn't belong to } S \end{array} \right\}$

В методе, предложенном А. Al Marashli и О. Al Dakkak, использованы две модели для данных формул: λ_s – модель голоса гипотетического говорящего, находящаяся в уже построенных моделях, λ_{back} – модель голосов, состоящая из всех остальных моделей голосов говорящих [5, с. 1].

Используя рассмотренный алгоритм, представляется возможным построить программу, фильтрующую голосовые потоки в соответствии с заданным вектором кепстральных коэффициентов.

Библиографический список

1. А. Al Marashli, О. Al Dakkak Automatic, Text-Independent, Speaker Identification and Verification System Using Mel Cepstrum and GMM / Infor-

mation and Communication Technologies Conference: “From Theory to Applications”, May 2008. 23 pages.

2. L.R. Rabiner, M.R. Sambur An Algorithm for Determining the End-points for Isolated Utterances / The Bell System Technical Journal, Vol. 54, N.2, Feb 1975, pp.297-315

3. F. Bimbot A Tutorial on Text-Independent Speaker Verification / EURASIP Journal on Applied Signal Processing 2004:4, pp.430–451

4. D. A. Reynolds, T. F. Quatieri, R. B. Dunn Speaker Verification Using Adapted Gaussian Mixture Models / M.I.T. Lincoln Laboratory, Digital Signal Processing 10, 19–41 (2000)

5. P. P. Singh, P. Rani An Approach to Extract Feature using MFCC /IOSR Journal of Engineering (IOSRJEN) Vol. 04, Issue 08 (August. 2014), ||V1|| PP 21-25

6. Аль Мандили М.Р.А., Лебедев В.В., Чернышев О.Л. Автоматизация процедуры голосовой аутентификации пользователей // Информационные ресурсы и системы в экономике, науке и образовании: сборник статей IX Междунар. научно-практ. конф. / под ред.А.П. Ремонтова. Пенза, 2019. С. 7-11.

**Аль Мандили
Мухамад Руслан Ахмед**
Тверской государственный
технический университет,
г. Тверь, Россия

Al Mandili Muhamad R.A.
Tver State Technical University,
Tver, Russia

Лебедев Владимир Владимирович
Тверской государственный
технический университет,
г. Тверь, Россия

Lebedev V.V.
Tver State Technical University,
Tver, Russia

Чернышев Олег Леонидович
Тверской государственный
технический университет,
г. Тверь, Россия

Chernyshev O.L.
Tver State Technical University,
Tver, Russia