EVALUATION CRITERIA FOR A VOICE IDENTIFICATION ALGORITHM BASED ON A THRESHOLD QUADRATIC DISCRIMINANT FUNCTION

Urinboev Johongir

Digital technologies and artificial intelligence development research institute
E-mail: jahongir8010@gmail.com
DOI: 10.61587/mmit.tiue.uz.v1i1.357

Abstract. This paper examines the evaluation criteria for the effectiveness of an algorithm based on a quadratic threshold discriminant function used in voice-based speaker recognition. In the model, an acceptance interval is determined for each speaker based on a single informative feature extracted from the voice signal, and the tested voice sample is either accepted or rejected according to its correspondence to this interval. The paper presents the formulas for the overall accuracy of the algorithm, the false acceptance rate, the false rejection rate, and the overall error rate. In addition, the influence of the parameter λ, which determines the acceptance interval, on the FAR and FRR indicators is analyzed.

Keywords: *voice identification, speaker recognition, quadratic threshold function, accuracy, λ parameter*

Introduction

Speaker recognition by voice is considered one of the important areas of biometric identification. A voice signal reflects the individual structure of the human speech apparatus, pronunciation style, rhythm, and spectral features. Therefore, voice-based speaker identification systems are widely used in information security, mobile devices, remote authentication, access control, and forensics [1].

One of the important issues in such systems is not only speaker recognition itself, but also the evaluation of the reliability of the recognition result. This is because two main types of errors may occur in a voice authentication system: falsely accepting an impostor speaker and falsely rejecting a genuine user. The first case poses a threat to security, while the second case reduces the usability of the system. In this work, the voice-based speaker recognition algorithm is briefly described on the basis of a quadratic threshold discriminant function. The main focus is placed on analyzing the evaluation criteria of this algorithm and the influence of the λ parameter, which controls the acceptance interval, on the system results.

Research Methodology

Model based on a quadratic threshold function. In the voice identification process, first, one informative feature is extracted from the voice sample S_u:

$$x_i = \varphi(S_u)$$

where φ is the feature extraction function [2]. As a feature, the fundamental frequency, energy, formant, or one of the MFCC coefficients may be selected. In this work, the feature extraction process is not considered in detail, since the main focus is placed on evaluation criteria. For each speaker class K_j, the mean value and standard deviation of the feature are calculated based on the training samples:

$$\mu_j = \frac{1}{m}\sum_{r=1}^{m} x_{jr},$$

$$\sigma_j = \sqrt{\frac{1}{m-1}\sum_{r=1}^{m}(x_{jr} - \mu_j)^2}.$$

Then, the acceptance interval is defined as follows:

$$x_{j1}^{(b)} = \mu_j - \lambda\sigma_j \; x_{j2}^{(b)} = \mu_j + \lambda\sigma_j$$

where λ is a parameter that controls the width of the acceptance interval. The quadratic threshold discriminant function is given as follows:

$$D_j(x_i) = -\left(x_i - x_{j1}^{(b)}\right)\left(x_i - x_{j2}^{(b)}\right)$$

or in the general form:

$$D_j(S_u) = a_{2j}x_i^2 + a_{1j}x_i + a_{j0}.$$

Decision rule:

$$\begin{cases} D_j(S_u) > 0, & S_u \in K_j, \\ D_j(S_u) \le 0, & S_u \notin K_j. \end{cases}$$

If there are several speakers in the system, the maximum discriminant value is selected:

$$j^* = \arg\max_j D_j(S_u).$$

Decision rule:

$$Result(S_u) = \begin{cases} K_{j^*}, & D_{j^*}(S_u) > 0, \\ Unknown, & D_{j^*}(S_u) \le 0. \end{cases}$$

is expressed in this form [3].

The need to evaluate algorithm results. After developing a voice-based speaker recognition algorithm, it is important to evaluate its effectiveness. This is because the real capabilities of the algorithm are determined not only by the mathematical model, but also by the error rates obtained from test results. In a voice identification system, the following cases may occur:

1. A genuine speaker is correctly recognized.
2. A genuine speaker is falsely rejected.
3. An impostor speaker is correctly rejected.
4. An impostor speaker is falsely accepted.

Based on these cases, the accuracy, security level, and user convenience of the algorithm are evaluated [4].

The influence of the λ parameter on the algorithm. In the proposed quadratic threshold model, the λ parameter is the most important adjustable parameter. This is because it determines the width of the acceptance interval for each speaker:

$$\mu_j - \lambda\sigma_j < x_i < \mu_j + \lambda\sigma_j.$$

If the value of λ is selected to be small, the acceptance region becomes narrow. In this case, the rejection of impostor users is strengthened, that is, FAR decreases. However, some voices of genuine users may also fall outside the acceptance interval. As a result, FRR increases.

Conversely, if the value of λ is selected to be large, the acceptance region expands. In this case, the probability of accepting a genuine user increases, that is, FRR decreases. However, the probability that impostor voices may also fall within the acceptance interval increases [8]. As a result, FAR increases. Thus, the influence of the λ parameter is summarized as follows:

$$\lambda \downarrow \Rightarrow FAR \downarrow, FRR \uparrow \quad \lambda \uparrow \Rightarrow FAR \uparrow, FRR \downarrow$$

Graphical interpretation of the acceptance region with respect to λ. In the following form, it can be observed that as the value of λ increases, the acceptance interval expands:

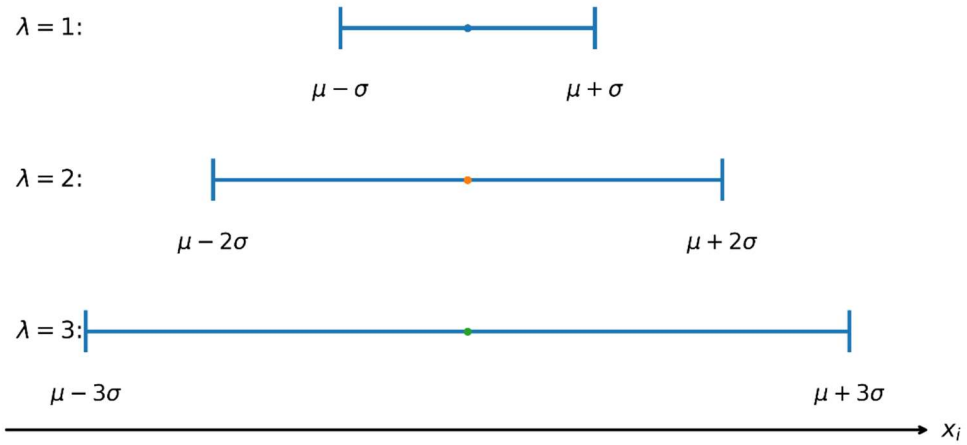

Figure 1. Selection of features according to the $\mu \mp \lambda\sigma$ criterion.

It can be seen from this graph that when $\lambda = 1$, the system operates more strictly. When $\lambda = 3$, the acceptance region becomes much wider. From the viewpoint of the quadratic function, an increase in λ expands the distance between the roots of the parabola:

$$x_{j2}^{(b)} - x_{j1}^{(b)} = 2\lambda\sigma_j.$$

Thus, the length of the acceptance interval directly depends on λ:

$$L_j = 2\lambda\sigma_j$$

where L_j is the length of the acceptance interval for the j-th speaker.

Selection of the λ *parameter.* In practical systems, the value of λ is selected based on experimental results. The optimal value should provide a balance between FAR and FRR.

Balance between FAR and FRR. In voice authentication systems, FAR and FRR have opposite characteristics. Narrowing the acceptance interval helps reject impostor speakers, but it also increases the probability of rejecting a genuine user. Expanding the acceptance interval makes it easier to accept a genuine user, but it increases the probability of falsely accepting impostor voices [9].

Conclusion

In this work, the evaluation criteria for a voice-based speaker recognition algorithm based on a quadratic threshold discriminant function were considered. The model was briefly described, and the main focus was placed on the accuracy, FAR, FRR, and overall error indicators used to evaluate the results. In the proposed algorithm, the acceptance interval for each speaker is defined by the following formula:

$$\mu_j - \lambda\sigma_j < x_i < \mu_j + \lambda\sigma_j$$

If the value of the quadratic discriminant function is positive, the voice sample is accepted as belonging to the corresponding speaker class; otherwise, it is rejected. In the evaluation process, the importance of the λ parameter was demonstrated. If λ is small, the acceptance region narrows and FAR decreases, but FRR increases. If λ is large, the acceptance region expands and FRR decreases, but FAR increases. Therefore, selecting the λ parameter is one of the main factors determining the effectiveness of the algorithm. In practical systems, the optimal value of λ should be selected based on test experiments, taking into account the balance between security and user convenience.

References

1. Rabiner, L. R., & Schafer, R. W. (2007). Introduction to digital speech processing. Foundations and Trends in Signal Processing, 1(1–2), 1–194. https://doi.org/10.1561/2000000001https://doi.org/10.1561/2000000001

2. Beigi, H. (2011). Fundamentals of speaker recognition. Springer. https://doi.org/10.1007/978-0-387-77592-0

3. Campbell, J. P., Jr. (1997). Speaker recognition: A tutorial. Proceedings of the IEEE, 85(9), 1437–1462. https://doi.org/10.1109/5.628714

4. Reynolds, D. A. (2002). An overview of automatic speaker recognition technology. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002) (Vol. 4, pp. 4072–4075). IEEE.

5. Fukunaga, K. (1990). Introduction to statistical pattern recognition (2nd ed.). Academic Press.

6. Theodoridis, S., & Koutroumbas, K. (2008). Pattern recognition (4th ed.). Academic Press.

7. McLachlan, G. J. (2004). Discriminant analysis and statistical pattern recognition. Wiley-Interscience.

8. Webb, A. R., & Copsey, K. D. (2011). Statistical pattern recognition (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781119952954

9. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.