Age and Gender Classification for Permission Control of Mobile Devices in Tracking Systems

Merahi Choukri(✉) and Shaochuan Wu

Harbin Institute of Technology, Harbin, China
merahichoukri1991@yahoo.com, scwu@hit.edu.cn

Abstract. Not only does the human voice provide the semantics of the spoken words but also it contains the speaker-dependent characteristics, such as the gender, the age, and the emotional state of the speaker. In the last decade, speech recognition gained a great interest in identifying and tracking systems. According to the speech length of ten to thirty seconds, this paper proposes an age and gender classification method for permission control of mobile devices. Each speech signal is firstly extracted to 40 features by Mel Frequency cepstral Coefficients (MFCC). After that, the Support Vector Machine (SVM) is used to finish the age and gender classification. This paper studies six kernel models of SVM and concludes that cubic, quadratic, and medium Gaussian kernel models could improve the recognition rate up to 93.75%, 91.25% and 93.75% respectively. Therefore, it is promising for permission control of a mobile in tracking systems.

Keywords: Classification · Permission control ·
Mel Frequency cepstral Coefficients (MFCC) · Support Vector Machine (SVM)

1 Introduction

For permission control of mobile devices based on speech recognition, several features must be extracted from humans' voices, such as the gender, the age, and the emotional state of the speaker. Although currently there are some available approaches to identify certain information about callers or speakers and classify them according to their emotion [1–3], age [4, 5] and gender [6, 7], the performance of classification is still need to be improved.

Support Vector Machine (SVM) is one of the most popular machine learning algorithms which have been widely used for the recognition of speech. Controlling the error level and improving the efficiency of speech recognition is still a big challenge. Several speech recognition methods have been proposed based on the SVM approach [8, 9]. Depending on the models used in SVM, the recognition accuracy changes in the range from 60% to 75% [10] which it too low to be used in permission control of mobiles.

In this work, the SVM approach with six kernel models (linear, cubic, quadratic, medium Gaussian, fine Gaussian and coarse Gaussian) has been used. As our best knowledge, cubic, quadratic, and medium Gaussian kernel models have not been used in previous literatures. Three main steps are adopted in this work. The first step is the signal preprocessing at which the noisy speech is preprocessed into noise-free speech.

S. Han et al. (Eds.): AICON 2019, LNICST 287, pp. 318–324, 2019.
https://doi.org/10.1007/978-3-030-22971-9_28

At the same time, the silence epoch of speech will be removed by voice activity detection. The second step is the speech features extraction realized by Mel frequency cepstral coefficients method. The final step is the age and gender classification based on features by the SVM classifier. By simulation analysis, the recognition rate of cubic, quadratic, and medium Gaussian kernel models is 93.75%, 91.25% and 93.75% respectively, which is obviously high than previous methods. Since the recognition rate is higher than 90%, our methods can be efficiently used for the permission control of mobile devices in tracking systems.

2 Methods

2.1 Database

This work was based on ELSDSR database, developed by the Department of Informatics and Mathematical Modeling (IMM) at Technical University of Denmark [11]. In this database, 22 speakers of different ages, genders (12 males and 10 females), and nationalities (no native speakers) were selected randomly. It is worth mentioning that there was no rehearsal when creating the database. Each speaker read seven sentences in training and two sentences in testing. The duration of each speech was between 10 and 30 s. The age structure was balanced between 20 years and 60 years with different labels for training and testing. The different speakers have been classified as follow:

- Class one (Medium female): age 20–39.
- Class two (Old female): age 40–60.
- Class three (Medium male): age 20–39.
- Class four (Old male): age 40–60.

2.2 Feature Extraction

The feature extraction method, Mel Frequency cepstral Coefficients (MFCC), was introduced in the 1980s by Davis and Mermelstein [12]. They considered a time window of 30 ms with a time shift of 10 ms and 40 coefficients. Based on this method, we can obtain a vector with 40 components to be further used in the classification procedure. MFCC is a well-known feature extraction method used to recognize speech, speakers and even emotion. One of the most vital advantages of MFCCs is that it is a more effective method in terms of noise and spectral estimation errors compared to other methods [14].

2.3 Algorithm

In this part, an effective age and gender recognition scheme will be introduced. In order to obtain better results, some pre-processing techniques were applied to the training data. The first pre-processing method was spectral subtraction method, which was applied to remove the background noise such as white noise or musical noise. After spectral subtraction method, voice activity detection was used to training data in order

to remove the silence epoch in the speech [13]. Due to its reliability, the MFCC was used as a potential spectral subtraction method for this scheme. At the end, we used SVM for training and testing (Figs. 1 and 2).

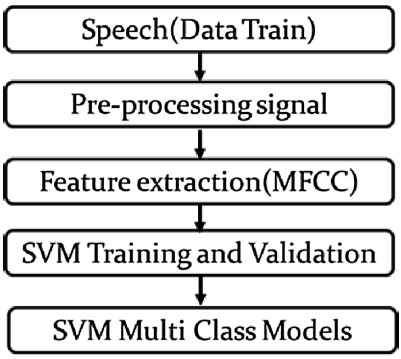

Fig. 1. Process of the training data.

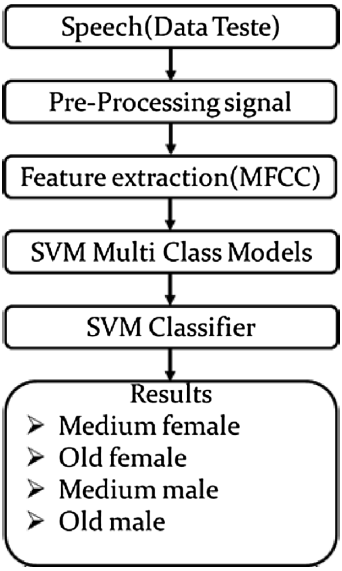

Fig. 2. Process of the testing data.

2.4 Training

Age and gender recognition is a multi-class classification task. After the previous pre-processing and MFCC feature extraction steps, an SVM was used for training both genders and age groups. Six kernel models were used for comparison. The four generated classes are already mentioned at Sect. 2.1. The training labels included medium

males and medium females whose ages range between 20 and 39 years old, and also old males and old females whose ages range between 40 and 60 years old for SVM training. The training was selected by cross-validation, and the SVM parameters were selected randomly.

3 Support Vector Machines

3.1 SVM Classification

The optimization objective function of Support Vector Machines can be described as Below [8]:

$$\min_\theta c \sum_{i=1}^{m} \left[y^{(i)} \cos t_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \cos t_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2 \tag{1}$$

where the idea is to try to minimize this objective function. If the parameter vector θ transpose times x is greater or equal than zero, it will be classified as 1 and otherwise, it will be classified as zero:

$$h_\theta(x) = \begin{cases} 1, & if\ \theta^T x \geq 0 \\ 0, & otherwise \end{cases}$$

A binary classification y ∈ {−1, 1} based on hyperplane separation was performed by Support Vector Machine SVM. In order to maximize the distance between the hyperplane and the closest training vectors, a support vector was chosen using the Kernel functions $K(x_i, x_j)$, these functions must satisfy the Mercer condition. As a result, the SVM can be extended to non-linear boundaries. Where y_i and x_i are target values and the support vectors respectively. λ_i must be determined during the training process. L represents the number of support vectors, and d is a (learned) constant:

$$f(x) = \sum_{i=1}^{L} \lambda_i y_i K(x, x_i) + d \tag{2}$$

The aim of our work is four-class identification, which allows us to extend the binary SVM. In order to extend the binary SVM, the simplest way is to take each class as an independent class. Therefore, a classifier has to be created for each class. The class of the speaker is determined according to the largest score of the classifier:

$$\arg\max \frac{1}{N} \sum_{K=1}^{N} \left(\sum_{i=1}^{L} \lambda_i y_i K(x, x_i) + d \right) \tag{3}$$

With j ∈ {1,…,N}.

3.2 SVM Models

Trying different models on the training and testing data for comparison will be very benefiting for future works, i.e. the optimized models can be used in future works in order to get optimum results. SVM has several types of kernel models, each model gives different results of training and classification, Herein, six kernel models were carried out, the results and the comparison between them will be introduced in the sequel.

4 Results and Discussion

In this paper, age and gender recognition was carried out in six experiments. Each experiment contains two steps to use ELSDER database. In the first step, the training was performed with all SVM models in order to get the best-trained models of SVM. In the second step, we performed testing with the trained models in order to get significant result of classification, which allows us to conclude the best model.

The computation of SVM models with training and classification are illustrated in Table 1.

Table 1. Training and classification of the database ELSEDER using six SVM models.

SVM models	Class 1	Class 2	Class 3	Class 4	Overall accuracy
Linear	40%	80%	100%	00%	55%
Cubic	100%	100%	100%	75%	93.75%
Quadratic	90%	100%	100%	75%	91.25%
Medium Gaussian	100%	100%	100%	75%	93.75%
Fine Gaussian	00%	00%	100%	00%	25%
Coarse Gaussian	70%	100%	100%	00%	68.75%

According to Table 1, the overall accuracy varied between 25% and 93.75%. The best results were given by the cubic model and medium Gaussian model (93.75%), in these two models three classes had a successful recognition rate of 100% for testing, and the fourth class had a successful recognition rate of 75%, which also can be considered acceptable.

The quadratic model also gave a significant overall accuracy (91.25%), two classes had a 100% successful recognition rate, class one had 90%, and class four had 75%.

The linear and coarse Gaussian models obtained poor testing results (55% and 68.75% successful recognition rate, respectively). Fine Gaussian model is the worst one with successful recognition rate 25%.

By comparison with polynomial and radial basis function kernel models proposed in [10], the cubic, the medium Gaussian and the quadratic models utilized in this paper are better for age and gender recognition.

5 Conclusion

This work performed an age and gender recognition and classification according to speech signals for permission control of mobile. The SVM approach based on six kernel models was carried out. By speech signal preprocessing technique, speech feature extraction and classification, we found that three models (the cubic model, the medium Gaussian model and the quadratic model) achieved overwhelming performance advantages. With above 90% successful recognition ratio, the new classification method can be useful for the permission control of mobile devices.

Acknowledgements. This work was supported by the National Key R&D Program of China under grant 2018YFC0806803.

References

1. Patel, P., et al.: Emotion recognition from speech with Gaussian mixture models & via boosted GMM. Int. J. Res. Sci. Eng. **3** (2017)
2. Adeyanju, I.A., Omidiora, E.O., Oyedokun, O.F.: Performance evaluation of different support vector machine kernels for face emotion recognition. In: SAI Intelligent Systems Conference (IntelliSys). IEEE (2015)
3. Sokolov, D., Patkin, M.: Real-time emotion recognition on mobile devices. In: 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018). IEEE (2018)
4. Qawaqneh, Z., Mallouh, A.A., Barkana, B.D.: Age and gender classification from speech and face images by jointly fine-tuned deep neural networks. Expert Syst. Appl. **85**, 76–86 (2017)
5. Wang, S., Tao, D., Yang, J.: Relative attribute SVM+ learning for age estimation. IEEE trans. Cybern. **46**(3), 827–839 (2016)
6. Akbulut, Y., Şengür, A., Ekici, S.: Gender recognition from face images with deep learning. In: 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE (2017)
7. Kaya, K., et al.: Emotion, age, and gender classification in children's speech by humans and machines. Comput. Speech Lang. **46**, 268–283 (2017)
8. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM trans. Intell. Syst. Technol. (TIST) **2**(3), 27 (2011)
9. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data min. Knowl. Discov. **2**(2), 121–167 (1998)
10. Bocklet, T., et al.: Age and gender recognition for telephone applications based on GMM supervectors and support vector machines. In: ICASSP (2008)
11. Feng, L.: Speaker recognition, informatics and mathematical modelling. Technical University of Denmark (2004)

12. Davis, S., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE trans. Acoust. Speech Signal Process. **28**(4), 357–366 (1980)
13. Sohn, J., Kim, N.S., Sung, W.: A statistical model-based voice activity detection. IEEE Signal Process. Lett. **6**(1), 1–3 (1999)
14. Kim, H.-J., Bae, K., Yoon, H.-S.: Age and gender classification for a home-robot service. In: The 16th IEEE International Symposium on Robot and Human interactive Communication, RO-MAN 2007. IEEE (2007)