Classification Influence of Features on Given Emotions and Its Application in Feature Selection

Yin Xing¹, a, Chuang Chen², b and Li-Long Liu¹, c
¹College of Geomatics and Geoinformation, Guilin University of Technology, Guilin, China
²College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing, China

¹xycc123@foxmail.com; bchenchuang@njtech.edu.cn; chn_liulilong@163.com

Abstract. In order to solve the problem that there is a large amount of redundant data in high-dimensional speech emotion features, we analyze deeply the extracted speech emotion features and select better features. Firstly, a given emotion is classified by each feature. Secondly, the recognition rate is ranked in descending order. Then, the optimal threshold of features is determined by rate criterion. Finally, the better features are obtained. When applied in Berlin and Chinese emotional data set, the experimental results show that the feature selection method outperforms the other traditional methods.

1. Introduction

Speech emotion recognition is a research hotspot in the fields of artificial intelligence and signal processing. Its purpose is to enable the computer to perceive the surrounding environment, atmosphere and emotional state of the object as human beings, so as to provide the best conversation environment for dialogue objects [1]. The findings of Stanford University research by Reeves and Nass [2] show that the problems that need to be solved in human-computer interaction are in fact consistent with the important factors of human-human interaction. The most crucial is the ability of "emotional intelligence". Therefore, if the computer is able to more actively adapt to the needs of the operator, it must first identify the operator's emotion, and then adjust the way of interactive dialogue according to emotional judgment. The research has important application value in the fields of new human-computer interaction [3], call centre [4] and intelligent robot [5].

For speech emotion recognition system, it consists of three steps: acquisition of speech signal, extraction of emotion features and recognition of speech emotion [6]. At present, researchers have proposed many effective speech emotion features such as short-time energy, pitch frequency, formant, and Mel Frequency Cepstral Coefficient (MFCC). Although these features have achieved good results in speech emotion classification, the features often have higher dimensions. In particular, after the fusion of multiple types of features, the amount of feature data greatly increases. High-dimensional feature sets contain large amounts of noise and redundant data. On the one hand, it affects the recognition rate; on the other hand, it increases the computational complexity [7]. Therefore, it is necessary to perform feature reduction on the extracted features.

To sum up some literature in the field of speech emotion recognition, researchers mainly adopt the following methods of dimensionality reduction. Chiou et al. [8] used Principal Component Analysis (PCA) to process speech emotion features in Berlin emotional data set. After removing a lot of
features, it still got 80% recognition rate. Xue et al. [9] took into account the nonlinear pop structure of speech signals and used Locally Linear Embedding (LLE) to realize feature reduction. Yuan et al. [10] used improved Linear Discriminant Analysis (LDA) to carry out experiments on Chinese emotional data set and achieved better rate. Zhang et al. [11] used Fisher Discriminant Ratio (FDR) to select emotion features and got 81% rate in computer-induced emotional corpus. Although the above methods have achieved a certain reduction effect, they all lack the classification influence analysis of the features on given emotions. Thus, there is a certain limit on recognition rate. Therefore, it is of great importance to develop a new feature selection method.

2. Feature analysis and selection (FAS) method
Considering difference contributions of a certain feature on the given emotion, each feature need to be deeply analyzed. It can give the rank of feature importance that affects the rate of given emotion, so as to find out the key features in high-dimensional emotion features. In order to facilitate the narrative, this paper will the feature analysis and selection method abbreviated as FAS method. The operation process of FAS method is described in detail as follows:

1. Feature extraction. It is assumed that \( n \)-dimensional emotion features are extracted, forming the feature vector \( F = (f_1, f_2, \ldots, f_n) \).

2. Given an emotion for classification. Based on the speech emotion classifier, the recognition rate of the \( j \)-th emotion using the \( i \)-th feature \( f_i \) is \( r_{ij} \). Then the established recognition rate matrix \( R \) is as follows:

\[
R_{j \times m} = \begin{bmatrix}
    r_{11} & r_{12} & \cdots & r_{1j} & \cdots & r_{1m} \\
    r_{21} & r_{22} & \cdots & r_{2j} & \cdots & r_{2m} \\
    \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
    r_{1} & r_{2} & \cdots & r_{j} & \cdots & r_{m} \\
    \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
    r_{n1} & r_{n2} & \cdots & r_{nj} & \cdots & r_{nm}
\end{bmatrix}
\]  

(1)

3. Sort of feature importance. Each column of the matrix \( R \) is descended to obtain a sorted recognition rate matrix \( M \).

4. Determine the threshold \( k \). The best \( k \) features of each emotion in matrix \( M \) are selected to form a new feature set \( N \). The \( k \) value is chosen based on the actual recognition rate of speech emotion. That is, we think the \( k \) value is the best when maximum average recognition rate is found.

5. Streamline operations. There may be repeated features in the feature set \( N \), and only one feature of them is taken as a representative.

This process of selecting emotion features using FAS method is simple and easy to implement. It takes the actual recognition rate as selection criteria and has a good applicability.

3. Feature extraction and classification algorithm
The experimental data in this paper are derived from Berlin emotional speech database [12] and Chinese emotional speech database [13]. The Berlin emotional speech database was recorded by 10 (5 males and 5 females) non-professional actors. And a total of 800 emotional statements were obtained. These emotional statements include 7 emotions that are anger, happiness, neutral, sadness, fear, disgust and boredom. After 20 auditions, 535 sentences were retained. The Chinese emotional speech database was recorded by 4 professional speakers (2 males, 2 females). And a total of 1200 emotional statements were obtained. These emotional statements include 6 emotions that are anger, happiness, neutral, sadness, fear, and surprise. The speech samples of the two emotional speech databases are stored at 16000 sampling rate, 16bit quantization and wav format. In following
experiments, we selected 5 emotions as experimental data. The specific information is shown in Table 1.

| Database                          | Number of samples | Number of features |
|-----------------------------------|-------------------|--------------------|
| Berlin emotional speech database  | 126               | 124                |
| Chinese emotional speech database | 200               | 124                |

In this paper, 124 emotional features are extracted, which are maximum, minimum, mean and variance of short-term energy, pitch frequency, first-formant frequency, second-formant frequency, third formant frequency, MFCC and its first-order difference in 0 ~ 12th order. The calculation methods are detailed in reference [14]. Then, the K-Nearest Neighbor (KNN) algorithm is used for emotion classification. The basic idea of classification is that given a sample to be classified in feature space, if the majority of k nearest neighbors in a class belong to a certain class, then the sample to be classified currently also belongs to this class. In order to verify the reliability of FAS method, \( K = 3 \), \( K = 5 \) and \( K = 7 \) were used for classification experiment. Several experiments have found that the different values of \( K \) are only small changes in the recognition rate, but have no effect on the general trend of the recognition rate curve. Therefore, this paper selects \( K = 5 \) for the experiment.

4. Speech emotion classification experiment
First, the feature selection experiment was conducted on Berlin speech emotional database using FAS method. Given emotions are classified by each feature, shown in Table 2. Due to space limitations, only some of the results are given in Table 2. From Table 2, each feature has different contribution to different emotions. For example, the feature 1 has a greater contribution to anger, while feature 7 has a greater contribution to the fear.

| Feature | Anger  | Happiness | Neutral | Sadness | Fear  |
|---------|--------|-----------|---------|---------|-------|
| 1       | 0.4127 | 0.1176    | 0.1795  | 0.1290  | 0.3333|
| 2       | 0.4921 | 0.0294    | 0.2308  | 0.0645  | 0.3636|
| 3       | 0.3016 | 0.1471    | 0.1795  | 0.1935  | 0.2424|
| 4       | 0.2857 | 0.0294    | 0.2564  | 0.2258  | 0.2121|
| 5       | 0.5714 | 0.2353    | 0.2051  | 0.1935  | 0.3333|
| 6       | 1.0000 | 0.7647    | 0.0000  | 0.0000  | 0.0000|
| 7       | 0.2698 | 0.1176    | 0.1026  | 0.2258  | 0.3030|
| 8       | 0.3492 | 0.1471    | 0.2821  | 0.1935  | 0.2424|
| 9       | 0.4762 | 0.1765    | 0.2564  | 0.4194  | 0.5152|
| 10      | 0.3333 | 0.1471    | 0.3590  | 0.1935  | 0.5152|

The recognition rates of each emotion in Table 2 are arranged in descending order, and the feature importance ranking is obtained, shown in Table 3. It shows the best 10 features of each emotion. The experiment needs to select the best \( k \) features for the classification. For example, when \( k \) is 1, the feature set is \( \{6, 6, 99, 107, 119\} \). Since the feature 6 has been repeated, we do not repeat the selection, then the new feature set is \( N = \{6, 99, 107, 119\} \).

| Anger Rate | 1.0000  | Feature | 6        | 0.7647  | 6        | 0.4872  | 99       | 0.6774  | 107      | 0.6061  | 119                  |
| Happiness Rate | 0.8889 | Feature | 79       | 0.7059  | 75       | 0.4615  | 75       | 0.5161  | 15       | 0.5758  | 115                  |

| Neutral Rate | 0.7647  | Feature | 6        | 0.4872  | 99       | 0.6774  | 107      | 0.6061  | 119                  |
| Sadness Rate | 0.7059  | Feature | 75       | 0.4615  | 75       | 0.5161  | 15       | 0.5758  | 115                  |
| Fear Rate    | 0.4872  | Feature | 99       | 0.6774  | 107      | 0.6061  | 119                  |
Different $k$ values are selected to correspond to different feature sets. Figure 1 shows the average recognition rate curve of speech emotion under different $k$ values. From Figure 1, the average recognition rate of speech emotion first increases with the increase of $k$ value. When $k$ is 7, the average recognition rate reaches the maximum, which is 86.5%. When $k$ continues to increase, the recognition rate decreases. Therefore, 7 is the optimal threshold, and the number of corresponding feature set $N$ is 25.

![Figure 1. Average recognition rate curve of speech emotion under different $k$ values](image)

In order to verify the superiority of the FAS method proposed in this paper, we use PCA in [8], LLE in [9], LDA in [10] and FDR in [11] to reduce features. The experimental results are shown in Figure 2. From Figure 2, the recognition rate of FAS method proposed in this paper is obviously higher than that of the other 4 methods after reducing to 8 features, which shows the superiority.

![Figure 2. Recognition rate curves of different methods on Berlin speech emotional database](image)
In order to verify the reliability of the FAS method, this paper also conducted emotion recognition experiments on the Chinese speech emotional database. The PCA, LLE, LDA, FDR and FAS methods are used to reduce speech emotion features respectively, and KNN algorithm is used to classify speech emotion. The experimental results are shown in Figure 3. From Figure 3, the recognition rate curve of the FAS method is also higher than that of the other 4 methods. In summary, it is demonstrated that the FAS method proposed in this paper is effective and superior to the PCA, LLE, LDA and FDR methods.

![Figure 3](image_url)

**Figure 3.** Recognition rate curves of different methods on Chinese speech emotional database

5. Conclusion
This paper extracts a number of speech emotion features and analyzes the importance of features on given emotions. And the feature selection is carried out by FAS method. In the comprehensive comparison and analysis with the PCA, LLE, LDA and FDR methods, the experimental results verify that the FAS method is effective and improves speech emotion recognition rate.

Acknowledgments
This work is supported by the National Natural Science Foundation of China (41664002) and Guangxi Provincial Key Fund Project (2017GXNSFDA198016).

References
[1] L. Zhao. “Speech signal processing (third edition),” Beijing: China Machine Press, 2016, pp. 213-220.
[2] Y. M. Cheng, D. O'Shaughnessy, and P. Mermelstein. “Statistical Recovery of Wideband Speech from Narrowband Speech,” IEEE Transactions on Speech and Audio Processing, vol. 2, no. 4, 1994, 2(4), pp. 544-548.
[3] C. Jones and A. Deeming. “Affective Human-Robotic Interaction,” Affect and Emotion in Human-Computer Interaction. Springer-Verlag, 2008, pp. 175-185.
[4] D. Morrison, R. Wang, and L. C. De Silva. “Ensemble methods for spoken emotion recognition in call-centres,” Speech Communication, vol. 49, no. 2, 2007, pp. 98-112.
[5] R. W. Picard. “Robots with emotional intelligence,” Proceedings of ACM/IEEE International Conference on Human-Robot Interaction. IEEE, 2009, pp. 5-6.
[6] C. C. Huang, W. Gong, W. L. Fu, and D. Y. Feng. “Research of speech emotion recognition based on DBNs,” Journal of Computer Research and Development, no. s1, 2014, pp. 75-80.
[7] Y. S. Tao, K. X. Wang, J. Yang, and L. Li. “Hybridizing information gain and harmony search for feature selection on speech emotion,” Journal of Chinese Computer Systems, vol. 38, no. 5, 2017, pp. 1164-1168.
[8] B. C. Chiou and C. P. Chen. “Feature space dimension reduction in speech emotion recognition using support vector machine,” Proceedings of Signal and Information Processing Association Summit and Conference. IEEE, 2013, pp. 1-6.

[9] L. Xue and T. Qian. “Speech analysis based on locally linear embedding(LLE),” Proceedings of International Conference on Natural Computation. IEEE, 2010, pp. 2159-2162.

[10] J. J. Yuan, L. Chen, T. T. Fan, and J. Jia. “Dimension reduction of speech emotion feature based on weighted linear discriminant analysis,” International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 8, no. 11, 2015, pp. 299-308.

[11] X. D. Zhang, F. Hu, and L. Zhao. “Recognition of practical speech emotion based on improved shuffled frog leaping algorithm and support vector machine,” Signal Processing, vol. 27, no. 5, 2011, pp. 678-689.

[12] F. Burkhardt, M. Kienast, and A. Paeschke. “Berlin database of emotional speech (Technical University, Institute for Speech and Communication, Department of Communication Science, Berlin),” [2016-09-22]. http://pascal.kgw.tu-berlin.de/emodb/.

[13] Speech and human interaction research group, Institute of automation, Chinese Academy of Sciences. “Chinese database of emotional speech,” [2016-11-17]. http://www.datatang.com/data/39277.

[14] R. Y. Liang, L. Zhao, and X. Wei. “Experimental course on speech signal processing,” Beijing: China Machine Press, 2016, pp. 231-234.