International Journal of u- and e- Service, Science and Technology
Vol.8, No. 12 (2015), pp.217-226
http://dx.doi.org/10.14257/ijunesst.2015.8.12.22

An Improved Algorithm of Speech Emotion Recognition

Shulan Xia[1], Jilin Wang[2]*, Ru-gang Wang[2] and Li Zhao[3]

[1]*College of Electrical Engineering, Yancheng Institute of Technology, Yancheng 224051, Jiangsu, PR China*
[2]*College of Information Engineering, Yancheng Institute of Technology, Yancheng 224051, Jiangsu, PR China*
[3]*College of Information Science and Engineering, Southeast University, Nanjing 210096, Jiangsu, PR China*
xslnj@126.com

Abstract

The traditional K-nearest neighbor algorithm existences of the risk of miscarriage of justice, for its shortage a speech emotion recognition algorithm based on fuzzy K-nearest neighbors is proposed. By introducing the fuzzy membership concept, different characteristic parameters for the different contribution of emotion recognition are calculated, and the weighted Euclidean distance is used in speech emotion recognition. The experimental results show the effectiveness of the algorithm.

Keywords*: Speech emotion recognition ; fuzzy class membership function; FKNN (fuzzy K-Nearest Neighbor)*

1. Introduction

Nearest neighbor method is an important non parametric method; it takes all the training samples as the representatives, bypassing the probability estimation and calculation samples and sample to be measured and the distance to the nearest neighbor, the classes of the samples as the classification results. K nearest neighbor algorithm is extended to the nearest neighbor, the nearest neighbor algorithm from a sample of recent expansion into the nearest K samples, put up the sample categories appear as sample categories, but the K-nearest neighbor method also has some shortcomings, such as the need to consider all the samples stored in the decision risk and so on. In order to solve the limitation of the traditional K nearest neighbor method, in this paper, the K nearest neighbor algorithm is improved based on the fuzzy set theory, and the improved fuzzy K-nearest neighbor algorithm is applied in an emotional speech recognition.

2. K-nearest Neighbor Classification Algorithm and Fuzzy Set Theory

2.1. K Nearest Neighbor Classification Algorithm

K nearest neighbor (K-Nearest Neighbor, KNN) algorithm is a learning method based on statistical theory [1], and it is proposed by the Cover and Hart in 1968[1]. The present theory is relatively mature, and it is widely used in statistical machine learning. The basic idea is that in R^n space, for a sample to be classified, according to the distance, and K training of an unknown sample nearest neighbor sample points would be found, and then the samples to be classified category is determined according to the k neighbor sample points categories. For the problems of c class, a class $\omega_i (i = 1,2,...c)$ has known N_i training samples $x_j^{(i)}$ $(j = 1,2,...c)$. For a sample to be classified x, we respectively

ISSN: 2005-4246 IJUNESST

calculate the distance between the sample x and training sample $x_j^{(i)}$ with N_i. And the Euclidean distance is generally used to calculate. The K nearest neighbor sample, using majority voting, which kind of sample most, will x to determine which kind of, the principle of the algorithm is shown in Figure 1.

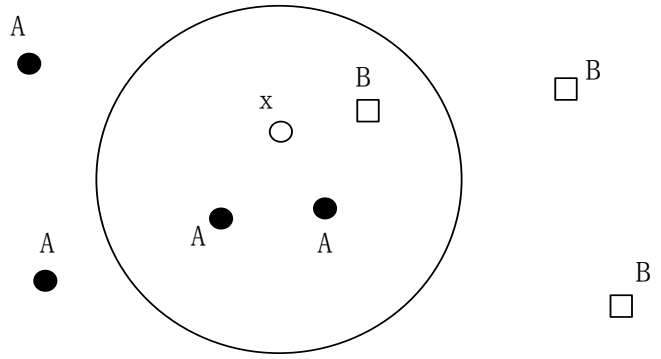

Figure 1. The Principle of KNN Algorithm

Implementation of K-nearest neighbor algorithm consists of two phases: (1) the training phase: discretization of the training sample, reading and storage; (2) the classification stage: firstly, the sample to be classified, the Euclidean distance is calculated between it and other training samples. By comparison of these distances, one of the nearest K samples is chosen. By studying the classification markings of the K neighbors, the sample category would to be determined according to the majority voting. Repeat the above steps until all the samples to be classified classification end.

K-nearest neighbor algorithm can not only be applied to classification, but also can be applied to the regression. By calculating the mean value properties of a sample of the K nearest neighbors, the properties of the sample can be estimated. K nearest neighbor algorithm is a simple and effective classification algorithm. And its advantage lies in is of simple, and easy to realize, the training process is very simple and fast. In the actual classification, sample categories associate with small amounts of nearest neighbor sample, it can better avoid the unbalance problem of sample. K-nearest neighbor algorithm does not rely on the identification class field to Classification, in the classification of the samples of the Cross of the class field or more overlap, using KNN algorithm can achieve better results. But the traditional K nearest neighbor algorithm also has some problem to be studied: (1) the selection of K value, the selection of K value is directly related to the algorithm of K nearest neighbor classification accuracy. Practice generally it require several experiments in order to find the optimum K value; (2) in the determination of the unknown sample categories assuming every known samples of the same role, it will appear in the known samples brought inconvenience caused by overlapping time, decision risk; (3) the large amount of calculation, before the tested samples are classified, calculate the distance of all training samples and it must be calculated, therefore K nearest neighbor algorithm is more suitable for classification in the case of small samples, the curse of dimensionality and K nearest neighbor algorithm may cause the dimension disaster in the high dimensional space. Therefore, how to improve the effectiveness of the K nearest neighbor algorithm is worthy of further in-depth study in high dimensional space.

2.2. Fuzzy Set Theory

The K-nearest neighbor algorithm is a simple and effective classification method, but there are some problems such as the determination of sample to be tested the hypothesis class when each sample does, although the unknown sample in which categories. It can't give the unknown sample belongs to the category of degree. In order to solve these problems, the fuzzy theory is introduced into the K-nearest neighbor classifier.

Fuzzy set theory is proposed in 1965 by USA cybernetics expert L.A. California university professor Zadeh [2-3], and it is widely investigated in recent years, which provides favorable mathematical tool for describing and studying the fuzzy phenomenon. Fuzzy set theory refers to the basic concept of the application of the fuzzy set in the theory or membership function. The basic idea is to accept the existence of fuzzy phenomena in the fact that the goal of the research is the thing the concept of fuzzy uncertainty, and the rigorous quantification for computer processing of information. The core of the theory is the object feature function concept in the ordinary set extended properties used to describe a concept of fuzzy membership function described as all the. In the application of the theory of pattern recognition and classification of fuzzy set, a sample of the membership to a certain degree of belonging to a certain class, rather than explicit is classified as belonging to a certain class or belonging to a class. Using a sample of all kinds of other membership function to describe the membership of the sample is more in line with the actual.

Fuzzy set theory is based on knowledge expression in real life, inductive learning and other aspects of the imprecise and incomplete, it can solve the problem of classical logic. In recent years, fuzzy set theory has been the rapid development, and varieties of development are produced in the form of description and solving all kinds of problems in the process of concrete.

Conclusion for domain U, any of its fuzzy subset of A, arbitrary $x \in U$, any of its fuzzy subset of A, arbitrary $x \in U$, there are a number of $u_A(x) \in [0,1]$, where $u_A(x)$ is the membership degree of x on A. Mapping $u_A(x): U \to [0,1]$, $x \to u_A(x)$, known as the fuzzy subset of the membership function, fuzzy subset is completely determined by its membership function defined by.

Suppose the set N has S samples, they belong to C category, a sample x_j for degree of membership categories i for u_{ij} , and u_{ij} should meet the conditions for:

$$u_{ij} \in [0,1] \tag{1}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \tag{2}$$

$$0 < \sum_{j=1}^{N} u_{ij} < N \quad . \tag{3}$$

3. Fuzzy K Nearest Neighbor Classification Algorithm

In 1985, fuzzy K-nearest neighbor algorithm is proposed by Keller on the basis of the existing theories to solve the limitation of the traditional K nearest neighbor algorithm. Due to the good robustness, the fuzzy K-nearest neighbor algorithm is widely used in machine learning. The basic idea of the fuzzy K-nearest neighbor algorithm and K nearest neighbor algorithm is the same, but the fuzzy K- nearest neighbor algorithm improvement of K nearest neighbor algorithm is to introduce the concept of category membership, rather like the K-nearest neighbor method is as simple as the sample only classified as a

class [4-6]. The membership degree is defined in samples belonging to different categories such as Eq.4.

$$u_i(x) = \frac{\sum_{i=1}^{k} u_i\left[x^{(j)}\right]\left[\left\|x - x^{(j)}\right\|^{-2/(m-1)}\right]}{\sum_{j=1}^{k}\left[\left\|x - x^{(j)}\right\|^{-2/(m-1)}\right]} \tag{4}$$

In the n-dimensional space R^n , for the C problem, the training sample set for $\{X_1, X_2, \ldots X_N\}$, N is the number of samples, for the sample of X tested, the specific algorithm steps of fuzzy K nearest neighbors are as follows:

(1) Determine the sample K neighbor K value;

(2) Determine the sample and all training samples to be measured the distance, the choice of using Euclidean distance:

$$d = d(x, X_i) = \sqrt{\left\|x - X_i\right\|^2} \tag{5}$$

(3) Sort the N distance：

$$d(1) \le d(2) \le d(3) \le \ldots \le d(k) \le d(k+1) \le \ldots \le d(N) \tag{6}$$

Where, $d(1), \ldots d(k)$ is the test samples and K nearest neighbor distance.

(4) According to the Eq. 1, calculate the sample category membership, where, m is a fuzzy weight adjustment factor. If $u_i(x) = \max\{u_n(x)\}$, then X is of the i class. The algorithm is repeated until all the tested samples are processed.

In recent years, fuzzy K nearest neighbor algorithm has been widely used in various involves pattern recognition field. For example, pattern recognition in Radix Puerariae, using fuzzy K nearest neighbor classification method, the correct recognition rate reaches 100% [7].

4. Speech Emotion Recognition based on FKNN

In this paper, the fuzzy K-nearest neighbor algorithm is used in speech emotion recognition research and its improvement on the recognition effect, mainly includes two aspects: extraction of speech emotion features and simulation experiment based on speech emotion recognition fuzzy K-nearest neighbor algorithm.

4.1. Speech Emotion Feature Extraction

Speech emotion features are selected in this paper for the short-time energy, short-time amplitude and short-time zero crossing rate and pitch frequency four parameters. Short-time energy, short-time magnitude, and the short-time zero crossing rate by a function Time Para ( ) to extract. The process is shown in Figure 2. The pitch frequency is extracted by the function Funfre ( ). This function is used in spectrum compression method, and the selection of window is Hanning window. After extracted the zero crossing rate and pitch frequency of short-time energy and short-time amplitude, the characteristic parameter would compose the feature vector, as the input of fuzzy K nearest neighbor classification algorithm.

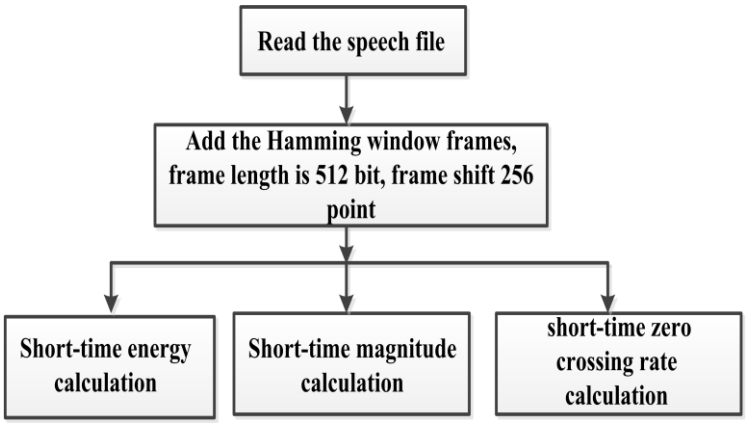

Figure 2. Flow Chart for Extracting the Short-time Energy, Short-time Amplitude, Short Time Crosses

4.2. Speech Emotion Recognition based on Fuzzy K -nearest Neighbor

In the training sample stage, K nearest neighbor classification method is different from the traditional, fuzzy K-nearest neighbor needed training samples are calculated for different categories of membership degree, mainly through the different emotional features for different categories of the contribution of the emotion recognition to achieve. Study a proof, in speech emotion recognition, emotional characteristics of different parameters for different emotions play a different role, for example, For example, for the short-time energy of this parameter, it can be very good to distinguish between neutral and sad these two emotions, the pitch frequency of neutral and sad two kind of emotion is similar, it is difficult to through the pitch to distinguish, but their energy has great difference, neutral energy is obviously larger than the sad energy. Therefore, in the distinction between neutral and sad emotions, energy contribution is larger than based audio. Furthermore you can distinguish between anger and happy that
the two kind of emotion through the pitch frequency, the energy of anger and joy are
relatively high, but the fundamental frequency was significantly higher than that of happy anger, so we can more accurately differentiate by calculating the different emotional characteristics of different feelings contribution to emotion [8].

In order to distinguish between different emotional features, first of all, emotional characteristic parameter must be statistics for the discrete degree of each kind of emotion, discrete degree is high, the emotion feature in distinguishing such feelings of uncertainty is high, and that is, for the emotion recognition of contribution degree is smaller. On the contrary, if the discrete degree is lower, the emotional feature parameter on the stability of this kind of emotion recognition is better, and the degree of contribution is the greater. The combination of the above thought, emotion feature contribution degree algorithms specific steps are as follows:

1) Identification of C kind of emotion, firstly, on the training sample set X, the average value with characteristic parameters has to be statistics of C different emotional state, that is M_{ij} ($i = 1, 2, \ldots C$, $j = 1, 2, \ldots N$), then each feature parameters M_{ijn} is normalized, the normalized formula is shown Eq. 7.

$$A_{ijn} = M_{ijn} \Big/ \sum_{i=1}^{c} M_{ij} \tag{7}$$

2) Calculate the dispersion of the characteristic parameters of a particular affection:

$$\theta_{ij} = \sqrt{\sum_{k=1}^{n} A_{ijk}} \tag{8}$$

3) After calculating each feature parameters of each kind of emotion of the dispersion, the contribution of each feature is calculated. The contribution u_{ij} can be expressed as:

$$u_{ij} = \theta_{ij} \Big/ \omega_i, \quad \omega_i = \sum_{l=1}^{j} \theta_{il} \tag{9}$$

And in the treatment of classification samples are distinguished using fuzzy K nearest neighbor, the emotion feature contribution degree is weighted with Euclidean distance.

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} d(x, X_j)}{\sum_{j=1}^{k} d(x, X_j)} \tag{10}$$

The emotion characteristic parameters of different emotion recognition degree of contribution as a weighted value, To be considered in the calculation, not only keeps the advantages of the K nearest neighbor algorithm is simple and easy to implement, but also highlight the characteristics of every emotion parameter plays a role in the differences, and the internal relation between different emotional features, improves the classification accuracy of K nearest neighbor algorithm, improvement of speech emotion recognition results.

5. Experiment of Speech Emotion Recognition based on FKNN

5.1. Experimental Environment and Emotional Speech Database

The experimental simulation platform is a PC with 2.6G/1G, Windows2003 XP /MATLAB7.0, and the voice toolbox of Voice box is also be used. The speech database for Chinese emotional speech database recorded on our own is selected. The sampling frequency is 16 KHz, the 16bit quantization. The speech database consists of ten recordings personnel (5 men and 5 women) in five different emotional state (happy (Happiness), anger (Anger), neutral (Neuter), sad (Sadness), fear (Fear)) under 20 different text reading word Chinese speech composition, after validation after a total of contains 624 effective emotional speech. The four kinds of emotions (happy, angry, the neutral, sad) have been selected. Emotional characteristic of short-time energy, short-time amplitude, short-time zero crossing rate and pitch frequency is selected. In order to reduce the impact of individual differences in emotional expression caused by the result of identification, speech emotion feature extraction to the normalized, and the emotional characteristics of the normalized as training samples and test samples in the experiment.

5.2. Emotional Feature Extraction

In the training phase for the four kind of emotion, the selected 30 sentences as a sample and feature parameter extraction of happy, angry, sad, neutral extracting short-time average amplitude energy, short time average energy, short-time zero crossing rate and pitch frequency [9]. It can be seen that in the neutral and sad pitch frequency is close to the short-time average energy, and short-term average energy have the obvious difference. But anger and joy of short-time average energy is closer to the pitch, and the pitch frequency angry is higher. These results show that the contribution and the Euclidean

distance using the feature parameters of fuzzy K nearest neighbor algorithm based on weighted for speech emotion recognition is reasonable.

After the characteristic parameters of the four emotions were extracted, according to the method obtained in 4.2, every characteristic parameter for the contribution of different emotional state recognition are calculated, shown in Table 1.

Table 1. Emotion Recognition Degree of Each Feature Parameters

Degree / Emotional state	Short-time Energy	average amplitude	Zero crossing rate	Pitch frequency
anger	0.5587	0.7324	0.9636	0.6396
sadness	0.6068	0.7460	0.9739	0.7144
happy	0.5863	0.7548	0.9697	0.6594
neutral	0.6237	0.7563	0.9534	0.7059

In the recognition stage, from the emotional statement four known were randomly selected 30 sentences in total of 120 word speech recognition and statistical recognition results, record and made numerical form of chart analysis and comparison.

5.3. Experimental Results and Analysis

The traditional K-nearest neighbor algorithm and fuzzy K nearest neighbor algorithm are adopted for pattern recognition. And the two methods are compared and the respective characteristics are analyzed through the experimental results. In the use of the traditional K nearest neighbor and fuzzy K nearest neighbor algorithm, the K has to select, in this paper, two series of experiments were carried out, the first group of experiments k=7, second groups of k=13. The recognition results are shown in Table 2 and table 3. It can be seen from the experiment results, the average recognition rate increase in k=13 than k=7. The four kind of emotion recognition rate were also having varying degrees of increase, this is mainly because of more neighbors, which reduces the risk of miscarriage of justice, but also increases the amount of computation. At the same time, it can be seen that because the sample size is small, the traditional K nearest neighbor algorithm and fuzzy K nearest neighbor algorithm recognition rate is relatively high. For the recognition of four emotions anger, sad, happy, neutral, recognition effect for anger is the best, mainly reason is that anger emotion characteristics is more obvious compared to the other three emotion. In expressing the emotion of anger, the speaker's speed is often faster, and the tone is high, for the neutral and sad recognition rate is relatively low, mainly because the sadness and neutral state, some physiological characteristics are similar, easily confused by mistake. From the comparison of two algorithms we can also find that fuzzy K nearest neighbor algorithm calculates the Euclidean distance to take full account of the proportion between various parameters compared with the traditional K nearest neighbor algorithm, which makes the recognition effect, is more prominent, and the recognition rate has improved in a way.

Table 2. The Recognition Results of Two Algorithms k=7

Identification method	Anger	Happy	Neutral	Sadness	The average recognition rate
KNN	82.53	81.28	76.47	78.14	79.61
FKNN	84.27	82.59	77.51	80.47	81.21

Table 3. The Recognition Results of Two Algorithms k=13

Identification method	Anger	Happy	Neutral	Sadness	The average recognition rate
KNN	83.29	82.36	78.53	79.62	80.95
FKNN	85.32	83.59	80.14	81.47	82.63

6. Conclusion

Aiming at the deficiency of the traditional k nearest neighbor algorithm cannot highlight the emotional characteristics of different parameters have different effects and based on the fuzzy set theory , we propose a speech emotion recognition algorithm based on fuzzy K nearest neighbor which introduced the concept of fuzzy membership, calculated the characteristic parameters for the different contribution of emotion recognition, and the weighted Euclidean distance used in speech emotion recognition. According to the Experiments of speech emotion recognition by Mandarin emotional speech database, The results shows that taking the weight ratio between various parameters into consideration, fuzzy K nearest neighbor improve the recognition rate, and can obtain better recognition effect compared with the traditional K nearest neighbor algorithm.

Acknowledgments

This work is supported by the Yancheng Institute of Technology Talents Project of China and the National Natural Science Foundation of China under Grant No. 60975017.

References

[1] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", IEEE Transactions on Acoustic, Speech and Signal Processing, ASSP, vol. 28, no. 4, **(1980)**, pp. 357-366.
[2] L. A. Zadeh, "Fuzzy sets.Information and control", vol. 8, no. 3, **(1965)**, pp. 338-353.
[3] H. Wang, "Nearest neighbors by neighborhood counting Pattern Analysis and machine Intelligence", IEEE Transaction, vol. 28, no. 6, **(2006)**, pp. 942-953.
[4] W. L. Gau, D. Buehrer, "Vague Sets", IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 2, **(1993)**, pp.610-614.
[5] K. Atanassov, "New Operations Defined over the Intuitionistic Fuzzy Sets", Fuzzy Sets and systems, vol. 61, no. 1, **(1994)**, pp.137-142.
[6] K. Atanassov, "Remarks on The Intuitionistic Fuzzy Sets-I II", Fuzzy Sets and Systems, vol. 75, no. 3, **(1995)**, pp. 401-402.
[7] M. Yang, "Study on the Pattern Recognition of Pueraria DC based on Fuzzy k-nearest Neighbor Algorithm", Chinese Journal of Biomedical Engineering, vol. 21, no. 5, **(2007)**, pp. 839-840.
[8] H. Zhu and J. B. Zhao, "Fuzzy K-nearest Neighbor Method and Its Application in Color Recognition", Computer CD Software and Applications, vol. 23, no. 3, **(2011)**, pp.165-166.
[9] J. L. Wang, S. L. Xia and Z. Li, "Speech Emotion Recognition of FSVW Based on FKNN", Microelectronics & Computer, vol. 31, no. 4, **(2013)**, pp. 31-35.

Authors

Xia Shulan, she received MS from Nanjing University of Science and Technology, Chian, in 2011. Now She is Associate professor in College of Electrical Engineering, Yancheng Institute of Technology, China. She current research interests include electronic technology and information processing.

Wang Jilin, he received MS from Southeast University, Chian, in 2004.Now he is full professor in College of Information Engineering, Yancheng Institute of Technology,China. His current research interests include signal and information processing.