International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.10 (2015), pp.57-66
http://dx.doi.org/10.14257/ijmue.2015.10.10.06

A Study on Speech Emotion Recognition Based On Fuzzy K Nearest Neighbor

Zhu Ming, Zhou Feng and Ji Zhengbiao

Yancheng Institute of Technology, College of Information Engineering, Yancheng 224051, China
zhumycit@163.com

Abstract

In order to improve the recognition performance of the tradition KNN, we combined the contributions of each feature parameter for different emotions with the tradition KNN and proposed a Fuzzy KNN algorithm for speech emotion recognition. Four kinds of emotions' recognition experiment has shown that the Fuzzy KNN we proposed not only keeps the tradition KNN's advantages of fast recognition and easy realization, but also improves the recognition performance.

Keywords: *Speech signal, Emotion feature analysis, Fuzzy KNN, Emotion recognition*

1. Introduction

Language is a very important tool for us to communicate, especially for hearing-impaired [1-2]. Our speech includes not only symbols and words but also our emotion and mood. For example, listener will feel different if speakers were in a different mood when speaking the same sentence. The traditional information science only deals with the "non-neural" knowledge world, such as the accuracy of the deliberation of information, but totally neglects the emotional element. So it only reflects one aspect of information. The emotional science world is corresponded to the knowledge science world, it's also an important component of information processing. So the artificial emotional feature processing is of very important meaning in the field of signal process and artificial intelligence [3-5].

KNN (K Nearest Neighbor) is a classical machine learning algorithm, and it is widely used with the advantage of easy implementation and fast recognition. Previous researchers have applied KNN in speech emotion recognition study and already achieved some good results. But as the traditional KNN can't consider the importance of different features, the recognition results are not always very ideal. In order to solve this problem, we analyzed four main emotion features first, then extracted 9 emotion feature parameters and proposed Fuzzy KNN algorithm to recognize speech emotion in this paper. Experiments on speech daTablease show that the Fuzzy KNN we proposed not only can get a higher recognition rate than tradition KNN, but also need little compute time and space. It has a high value of using.

2. Speech Emotion DaTablease

The choice of speech that fit for analysis is very important. But the standard and analytic condition of the speech for analysis has not been improved yet [6]. So this paper takes two aspects into account: First, the sentences don't have any emotional tendency. Second, the sentences must have a higher emotion freedom degree--it can involve all kinds of emotion to analysis and compare. According to the two rules, we use 60 sentences as the speech material. The classification of emotion is also a very important part of the research of emotional analysis. There're different methods to classify emotion

ISSN: 1975-0080 IJMUE

[7, 8, 14]. But in the view of engineering no subject has been proposed yet. In this paper, we roughly divided emotions into four types: happiness, anger, surprise and sadness. In order to obtain the original speech data, we had five male speakers who are good at acting to speak the 60 sentences one time in every emotion. Then, we had the speakers speak each sentence in a neutral way. In this way, we get 3000 sentences for experiment. Among them, 2000 sentences are used for training and the other 1000 sentences are used for recognition.

We recorded the speech daTablease in a quiet room by using Sony DAT equipment. The recorded data were transferred into digital signal by the PCI 64 bit sound card (made by Chuangtong Company) and stored in a PC computer. Then in order to validate all the emotional speech data, we played them randomly, and let some listeners (none of them are speakers) decide which type of emotion each sentence was. The listeners judge the emotional types all by their feeling. After repeatedly comparison and McNemar test, we selected the fittest sentences and reproduced the inconspicuous emotion sentences.

3. Emotional Feature Analysis

General speaking, the emotional feature of speech signal are always represented as the change of prosodic [15]. For example, when a man is angry, his speech rate, volume and tone will all get higher. People can feel these changes directly. However, as the emotional information of speech signal is more or less affected by the meaning of the sentence, usually we can find the relationship between emotional speech and non-emotional speech by analysis the construction features and distribution rules of speech characteristics to process and recognize different emotional speech signals. In this paper, we compared the feature of the time, amplitude, pitch and formant construction of four types of emotional speech to neutral ones to discover the distributing rules and construction features of different emotional speech. For convenience's sake, the ratios of the emotional feature parameters are shown below. They are the ratios between emotional feature parameters and neutral ones.

3.1. Time Construction Analysis

With an eye on the difference of different emotional speech's time construction, analysis of the time construction of emotional speech is to analyze and compare the difference of the change of duration arose by emotion. In this paper, we calculate the duration of each emotional sentence including the part of silence, because these parts contribute to the emotion. Then we compare and analyze the relationship between the average duration of emotional sentence (T) and the average rate of speaking with emotions (syllable/second). The results are shown in Figure 1.

From this Figure, we can see that the duration of anger and surprise is shorter than the one of neutral and happiness, but on the contrary, the duration of sad speech is longer. The duration of anger is shorter than the one of surprise. The duration of neutral speech is much shorter than the one of sadness and little shorter than the one of happiness. Through further observation, we can know these phenomena are caused by some faintly pronouncing, prolonged, omitted phoneme (compared to the neutral speech) in emotional speech. Based on the results above, we can recognize happiness, sadness and other emotion easily by comparing their duration to the neutral one. We can also set some time parameter threshold to recognize happiness and sadness. But we can't distinguish anger and surprise efficiently only by duration.

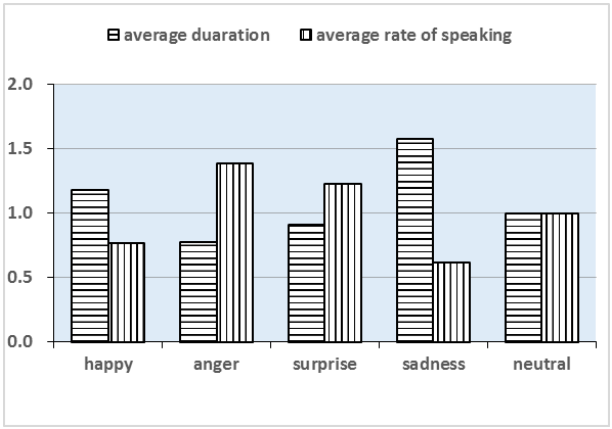

Figure 1. The Relative Values of Time Parameter of Various Emotions

3.2. Amplitude Construction Analysis

In general, the amplitude feature of speech signal closely relates to all kinds of emotional information [16]. We can feel in our real life that when a man is in a rage or in surprise, his volume is very high, however, when he is in sorrow, his volume is very low. So in some study of emotional analysis, amplitude construction is regarded as the most important feature. In our paper, we mainly take the average amplitude energy and its dynamic range into account (A and Arrange) to analyze and compare. We calculate the short-time energy of each frame of the signal, and analyze their characteristics which vary with time. To remove the effect of the silent and noisy part of the speech, we only take the average value of the absolute value of the amplitude into account. All the absolute values are bigger than a threshold. The results are shown in Figure 2.

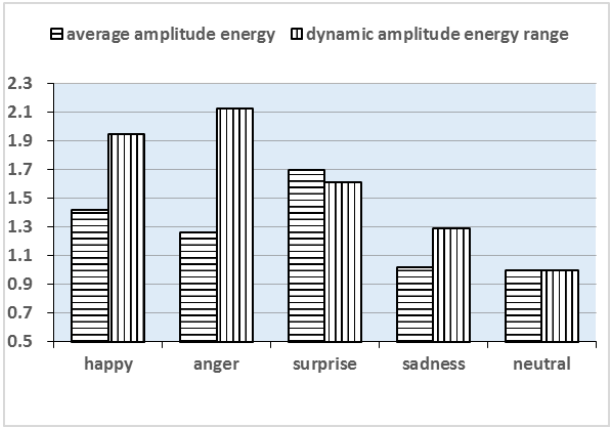

Figure 2. The Relative Values of Amplitude Parameter of Various Emotions

From the experiments results we can find that the emotional speech signals of happy, anger and surprise have bigger amplitude than that of neutral, while the amplitude of sadness is smaller than that of neutral. And we can also learn from the listening experiments that emotion signals have this trend, it is the bigger average amplitude of happy, anger, surprise have, and the smaller average amplitude of sadness have, the emotion effects are more obvious. So by using the feature of amplitude, we can easily distinguish happy, anger, surprise and sadness.

3.3. Fundamental Frequency Construction Analysis

Fundamental frequency construction is also an important feature that reflects emotional information [9]. To analysis the characteristic of fundamental frequency construction in emotional speech signal, we calculate the smooth fundamental frequency curves of the emotional speech signals, then analysis the characteristic of different fundamental frequency curves to find out the characteristic of the fundamental frequency constructions of different emotional speech signals. This paper analyzed the average fundamental frequency, its range, and its rate of change ($F0_{range}$, F0, $F0_{rate}$) of the curves of different emotional speech signals. Here $F0_{rate}$ refers to the mean absolute value of the difference between each frame of speech signal's fundamental frequencies. The results are shown in Figure 3.

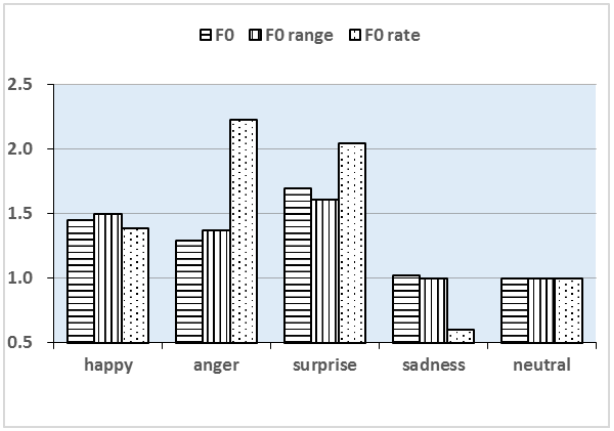

Figure 3. The Relative Values of F0 Parameter of Various Emotions

Compared to the neutral speech signal, the F0, $F0_{range}$ and $F0_{rate}$ of happiness, anger and surprise are bigger, while the ones of sadness are smaller. In view of happiness, anger and surprise, the parameters of surprise are the biggest, then are happiness and anger. In addition, we can see that the curve of surprising speech signal has the characteristic of raising at the end of the sentence. This characteristic is very useful for us to distinguish surprise from other emotion.

3.4. Fundamental Frequency Construction Analysis

Formant is an important parameter that reflects the characteristic of vocal cords [10]. We can predict the different location of formant of different emotional speech signals because different emotional speech signals change vocal cords differently. We get formant in two steps: 1) Apply LPC method to calculate the envelop of power spectrum of vocal cords; 2) Apply the Peak Picking method to calculate the frequency of the formant[12,13]. This paper only studies the average, range and rate of change of the first formant (F1, $F1_{range}$ and $F1_{rate}$). The results are shown in Figure 4.

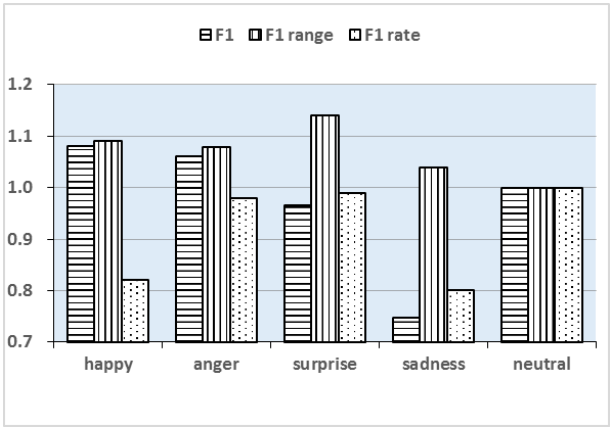

Figure 4. The Relative Values of F1 Parameter of Various Emotions

From Figure 4 we can see that compared to neutral, the frequencies of first formant of happiness and anger are a little higher while the ones of sadness are much lower. We can also learn by further observation that when we express the emotion of happiness or anger, our mouths always open bigger than usual and when we express emotion of sadness, our mouths open smaller and our speech always go with faint snuffle. The $F1_{range}$ of four types of emotion are bigger than the one of neutral. Among them, surprise is the biggest. However, the $F1_{rate}$ of all types of emotion are smaller than the one of neutral. Among them, sadness is the smallest.

3.5. Conclusion of Feature Analysis

From the above analysis of four speech emotion features, we can conclude the rules of speech emotion feature as shown in Table 1.

Table 1. The Changes of different Feature Parameters in Emotional Speech

	T	A	A_{range}	F_0	$F_{0\ range}$	$F_{0\ rate}$	F_1	$F_{1\ range}$	$F_{1\ rate}$
happy	+	+	++	+	+	+	+	+	_
anger	_	+	++	+	+	++	+	+	_
surprise	_	++	++	++	++	++	_	+	_
sadness	++	-	+	-	-	_ _	_ _	+	_ _

(The significance of symbols in the Table above. +:increase; ++:more increase; _:decrease; _ _:more decrease; -:no obvious change)

According to the analysis of speech emotion features, we chose 9 feature parameters for speech emotion recognition. They are the average duration of emotional sentence, the average amplitude energy and its dynamic range, the average fundamental frequency, its range, and its rate of change, the average, range and rate of change of the first formant. We extracted these 9 feature parameters from each emotion sentence to construct to one feature vector. As the units in the dimensions of feature vector are not the same, so before recognition, we normalized the elements in each dimension between 0 and 1. And after these steps, we used the feature vectors as the input of fuzzy KNN algorithm.

4. Speech Emotion Recognition Based on Fuzzy KNN

During training period, different from the traditional KNN, Fuzzy KNN needs to calculate the membership degree of training data for different classes [11, 17]. In this paper, we use the contribution of emotional feature parameters for different emotion

classes to represent the membership degree. Previous researches have shown that different emotional feature parameters plays a different role in different emotions' recognition. For example, short-time energy can distinguish neutral and sad very well, but they can hardly be distinguished by pitch frequency, because neutral and sad have the similar pitch frequency, meanwhile their energy have great difference. So short-time energy have more contributions than pitch frequency when distinguishing neutral and sad. However, we can use pitch frequency to recognition anger and happy. As anger and happy both have high energy, but obviously, anger has higher pitch frequency than happy. Hence we can have more accurate speech emotion recognition by calculate different emotion parameters' contributions for different emotions[11].

In order to distinguish different emotion parameters' contributions, firstly, we should count the dispersion of emotion parameters for each emotion. If the parameter gets higher dispersion, it means that this emotion parameter has more uncertainty in recognizing this emotion, and has less contribution for this emotion. Reversely, if the parameter gets lower dispersion, it shows this emotion parameter is more Tables when recognizing this emotion, and has more contribution.

Combining the above ideas, the recognition algorithm based on the contributions of emotional feature parameters is as follows:

1) For the recognition of C kinds of emotions, firstly, count the same feature parameter's mean value of C kinds of different emotions in training daTablease X, and recorded as $M_{ij}(i=1,2,...C, j=1,2,...N$,N stands for the number of emotion feature parameters). Then normalize the each feature parameter M_{ijn} (n stands for the index of example in one emotion, n=1 stands for the first sentence and so on) of each speech example in each emotion, the normalization equation is as follows:

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

2) Calculate the parameter's dispersion in one emotion:

$$\theta_{ij} = \sqrt{\sum_{k=1}^{n} A_{ijk}} \tag{2}$$

3) Calculate each parameter's contribution in each emotion according to the dispersion.

$$\omega_i = \sum_{l=1}^{j} \theta_{il} \tag{3}$$

The contribution u_{ij} of parameter θ_{ij} is:

$$u_{ij} = \theta_{ij} \Big/ \omega_i \tag{4}$$

Use the contribution as the Euclidean distance's weight when recognizing test examples by Fuzzy KNN classifier.

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} d\left(x, X_j\right)}{\sum_{j=1}^{k} d\left(x, X_j\right)} \tag{5}$$

Consider the contribution of emotion feature parameter for different emotion as the Euclidean distance's weight, not only keep the easy realization advantage of KNN, but also highlight the differences of each emotion feature parameter and their inner

relationships. By this means, we improved the classify accuracy of KNN and get higher speech emotion recognition effect.

5. Experiments

The simulation experiment platform is PC 2.6GHz/1GB, Windows 7 operation system/Matlab 2011b and speech tool software VoiceBox.

We choose four emotions (happy, anger, neutral, sad) in our speech emotion database to recognize. The feature parameters are mentioned in Chapter3. Before recognition experiments, we normalize the speech emotion features to reduce the feature parameters' individual differences. And the normalized emotion features were used as training examples and testing examples in following experiments. The training examples were total 2000 sentences, and they were selected from the speech emotion database randomly, and the rest 1000 sentences were used as testing examples.

During the recognition period, we used the traditional KNN and the Fuzzy KNN we proposed to recognize those testing examples. The extraction of 9 emotion feature parameters are as the above chapter3 shown. It is necessary to consider the size of k when using traditional KNN and Fuzzy KNN to recognize. We did two experiments, in the first experiment we chose k=7 and in the second experiment we chose k=13. The recognition results are as shown in Table 2 and Table 3.

Table 2. Recognition Results When k=7 k=7%

algorithm	anger	happy	neutral	sad	average recognition ratio
KNN	82.53	81.28	76.47	78.14	79.61
Fuzzy KNN	84.27	82.59	77.51	80.47	81.21

Table 3. Recognition Results When k=13 k=13%

algorithm	anger	happy	neutral	sad	average recognition ratio
KNN	83.29	82.36	78.53	79.62	80.85
Fuzzy KNN	85.32	83.59	80.14	81.47	82.63

From the experiment results, we can find that the average recognition ratio in k=13 experiment is higher than that in k=7 experiment, and the four emotions' respective recognition ratio is also improved. It is because more nearest neighbors were considered, and the risk of miscarriage was reduced, but more computation was needed.

Also, as the speech emotion database was not very large, both traditional KNN and Fuzzy KNN have got high recognition ratios. Among the experiment results of four emotions (anger, happy, neutral, sad), anger got the best recognition result. Because anger has more obvious emotion feature parameters than the other three kinds of emotions. When expressing anger emotion, speaker usually has faster speed and higher tone. While neutral and sad have similar psychological features, they are not easily to distinguish. From the comparative experiment of the two algorithms, we can find, as Fuzzy KNN takes a full consideration of the weight proportion of each parameter when calculating the Euclidean distance, Fuzzy KNN can get higher recognition results than the tradition KNN.

6. Conclusion

In this paper, we first construct a speech emotion database containing 3000 speech emotion sentences. Then the emotion feature analysis and extraction were discussed. In order to improve the recognition effect of traditional KNN, we proposed a Fuzzy KNN algorithm by combining the contributions of feature parameters for different emotions with the Euclidean distance. In this way, the Fuzzy KNN not only keep the fast, easily realization advantage of the traditional KNN, but also gets higher recognition ratio. Finally, experiments on 1000 test examples have shown the improvement and recognition effectiveness of the Fuzzy KNN we proposed.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (Grant No. 61273266, 61201326). The authors would like to thank the reviewers for their valuable suggestions and comments.

References

[1] Liang R., J. Xi and J. Zhou, "An improved method to enhance high-frequency speech intelligibility in noise", Applied Acoustics, vol. 74, no. 1, **(2013)**, pp. 71-78.
[2] Liang R.Y., J. Xi and L. Zhao, "Experimental study and improvement of frequency lowering algorithm in Chinese digital hearing aids", Acta Physica Sinica, vol. 61, no. 13, **(2012)**, pp. 1-13.
[3] Pereira M. G., L. de Oliveira, F. S. Erthal, "Emotion affects action: midcingulate cortex as a pivotal node of interaction between negative emotion and motor signals", Cognitive, Affective, & Behavioral Neuroscience, vol. 10, no. 1, **(2010)**, pp. 94-106.
[4] Huang C. W., R. Y. Liang and Q. Y. Wang, "Practical Speech Emotion Recognition Based on Online Learning: From Acted Data to Elicited Data", Mathematical Problems in Engineering, vol. 2013, **(2013)**, pp. 1-9.
[5] Huang C., G. Chen and H. Yu, "Speech emotion recognition under white noise", Archives of Acoustics, vol. 38, no. 4, **(2013)**, pp. 457-463.
[6] Jin Y., P. Song and W. M. Zheng, "Speaker-Independent Speech Emotion Recognition Based on Two-Layer Multiple Kernel Learning", Ieice Transactions on Information and Systems. E96D, no. 10, **(2013)**, pp. 2286-2289.
[7] Lee C. C., E. Mower and C. Busso, "Emotion recognition using a hierarchical binary decision tree approach", Speech Communication, vol. 53, no. 9-10, **(2011)**, pp. 1162-1171.
[8] Zhao X., E. Dellandrea and L. Chen, "Accurate land marking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 41, no. 5, **(2011)**, pp. 1417-1428.
[9] Jin Y., C. Huang and L. Zhao, "A semi-supervised learning algorithm based on modified self-training SVM", Journal of Computers, vol. 6, no. 7, **(2011)**, pp. 1438-1443.
[10] Jin Y., P. Song and W. Zheng, "Speaker-independent speech emotion recognition based on two-layer multiple kernel learning", IEICE Transactions on Information and Systems. E96-D, no. 10, **(2013)**, pp. 2286-2289.
[11] Xu, X. Z., C. W. Huang and C. Wu, "Graph Learning Based Speaker Independent Speech Emotion Recognition", Advances in Electrical and Computer Engineering, vol. 14, no. 2, **(2014)**, pp. 17-22.
[12] Zhang X., C. Huang and L. Zhao, "Recognition of practical speech emotion using improved shuffled frog leaping algorithm", Shengxue Xuebao/Acta Acustica, vol. 39, no. 2, **(2014)**, pp. 271-280.
[13] La V. T., V. P. Dao and X. Jim "Study on method of emotion recognition of speech based on simulated annealing genetic algorithm and support vector machine", International Journal of Advancements in Computing Technology, vol. 4, no. 20, **(2012)**, pp. 141-148.
[14] Tawari A. and M. Trivedi, "Speech emotion analysis in noisy real-world environment", 2010 20th International Conference on Pattern Recognition, ICPR, Istanbul, Turkey, August 23- 26, **(2010)**.
[15] Gong C., H. Zhao and Z. Tao, "Feature analysis on emotional Chinese whispered speech", 2010 International Conference on Information, Networking and Automation, ICINA, Kunming, China, October 17 - 19, **(2010)**.
[16] Thomaz C. E. and G. A. Giraldi, "A kernel maximum uncertainty discriminant analysis and its application to face recognition", 4th International Conference on Computer Vision Theory and Applications, VISAPP, Lisboa, Portugal, ISA, February 5-8, **(2009)**.
[17] Bondugula R., O. Duzlevski and D. Xu, "Profiles and Fuzzy K-Nearest Neighbor Algorithm for protein secondary structure prediction", 3rd Asia-Pacific Bioinformatics Conference, APBC, Singapore, Singapore, January 17- 21, **(2005)**.

Author

Zhu Ming, He was born in 1971 year. He is from Yancheng City. He is now working as a lecturer in information engineering school of Yancheng institute of technology. He has got the master's degree of information engineering School of Soochow University. Mr. Zhu Ming is providing efficient and high quality of speech signal processing research and electronic technology lessons.