World Applied Sciences Journal 14 (4): 616-625, 2011
ISSN 1818-4952

Emotion Recognition of Speech Using Small-Size Selected Feature Set and ANN-Based Classifiers: A Comparative Study

[1]*Mansour Sheikhan,* [1]*Mohammad Khadem Safdarkhani and* [2]*Davood Gharavian*

[1]Department Electrical Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran
[2]Department Electrical Engineering, Shahid Abbaspour University, Tehran, Iran

Abstract: In the recent years, emotion recognition of speech has noticeable applications in the speech processing systems. Most of the researches in this field have been focused on finding informative features and combining powerful classifiers that improve the performance of emotion recognition systems in different applications. In this paper, three feature ranking methods are used for emotion recognition from speech. These methods are Fisher score (FS), linear support vector machine (L-SVM) and mutual information (MI). For this purpose, a rich feature set with the size of 55 is used. Then, two distinct feature subsets with the size of 39 and 16 features are selected from the mentioned feature set. To investigate the performance of system with a small-size input feature set, eight high-ranked features are selected from each of the mentioned feature sets (with the size of 55, 39 and 16) and two types of neural networks (multi-layer perceptron (MLP) and radial basis function (RBF)) are used for emotion recognition. Experimental results show that using MI-based feature ranking method and MLP recognizer result in emotion recognition rates above 80% by employing a small-size feature set.

Key words: Emotion recognition · Speech · Feature ranking · Neural networks

INTRODUCTION

Humans transfer their emotional states to others through face or body movements and sufficient changes in their speech. Facial motion and the tone of speech play major roles in expressing emotions such as anger, hate, fear, happiness, sad, calm and boredom. All of the mentioned emotions give additional information to a listener. Certain emotional states are often correlated with the particular physiological states which in turn have quite mechanical and thus predictable effects on speech, especially on the pitch frequency (F_0), timing and voice quality [1]. In the recent years, emotion recognition from speech has noticeable applications in the speech processing systems [2-4], especially in speech recognition systems [5-9].

Most of the existing emotion recognition systems consist of three main functional blocks: Feature extraction, feature selection and emotion recognition. It is noted that the feature extraction is a critical functional block in an emotion recognition system [10].

Some factors such as the number and gender of speakers, dialect, age, language and skills are effective factors on the emotion recognition accuracy. Nowadays, most of the researches in this field have been focused on finding informative features and combining powerful classifiers that improve the performance of emotion recognition systems in different applications [11, 12].

Most of the traditional emotion recognition models have been based on the maximum likelihood Bayes (MLB) [13] and linear discriminate classification (LDC) [14]. In the recent decade, artificial neural networks (ANNs) [15-18], support vector machines (SVMs) [19-22], decision trees [1], K-nearest neighbor (KNN) [23, 24], Gaussian mixture models (GMMs) [25] and hidden Markov models (HMMs) [26, 27] have been used for emotion recognition.

Some examples of researches in the recent two decades are as follows: Dellaert *et al.* [13] have compared the performance of three classifiers: MLB, kernel regression and KNN in the recognition of sadness, anger, happiness and fear emotional states. They used the features that were based on the pitch contour and accuracy of 60 to 65% has been achieved. Lee *et al.* [28]

Corresponding Author: Dr. Mansour Sheikhan, P.O. Box: 11365/4435, Post-Graduate Center, South Tehran Branch, Islamic Azad University, Iran. Tel: +98 21 88605565, Fax: +98 21 33717140.

have used linear discrimination, KNN classifiers and SVM to distinguish two emotions: "negative" and "non-negative" emotions. They have achieved maximum accuracy of 75%. Petrushin [29] has developed a real-time emotion recognizer using neural networks for the call center applications and achieved classification accuracy of 77% for two emotions: agitation and calm. Yu *et al.* [30] have used SVMs for emotion recognition. They have developed classifiers for four emotions: anger, happiness, sadness and neutral. The average recognition accuracy of their proposed system was about 73%.

To reduce the size of features, the feature selection methods have been used in some researches. Considering the features at different levels such as frame-level, syllable-level and word-level and using them in an emotion recognition system has been reported in [22]. Some of the feature selection methods such as the sequential floating forward selection (SFFS) [15], the wrapper approach with forward selection [31], the forward feature selection (FFS), the backward feature selection (BFS) [23], the principal component analysis (PCA) and linear discriminate analysis (LDA) [32] have also been used for selecting features in the speech emotion recognition systems.

The effect of using a rich set of features including formant frequency-related, pitch frequency-related, energy and Mel-frequency cepstral coefficients (MFCCs) features on improving the performance of speech emotion recognition systems is investigated in this paper. To reduce the size of this rich feature set, C-Support feature vector classification is performed using SVM data classification ability. Then, three feature ranking methods (Fisher score (FS), linear SVM (L-SVM) and mutual information (MI)) are used. Finally, due to the success of ANNs in improving the performance of speech processing systems [4, 6, 33-38], we use two types of neural networks (multi-layer perceptron (MLP) and radial basis function (RBF)) for the emotion classification in this study.

The rest of this paper is organized as follows. In the next section, emotional speech corpus is introduced. C-support vector classification is reviewed in Section 3 with the aim of pre-classification of feature vectors. The FS, L-SVM and MI-based feature ranking methods are reviewed in Section 4 along with the feature ranking results in emotion recognition application. The proposed emotion recognition system is introduced in Section 5 along with the experimental results and comparisons with similar researches. Finally, the paper is concluded in Section 6.

Emotional Speech Corpus: Using 22 speakers, the emotional speech corpus has been recorded in this work. Each speaker has uttered 252 sentences in four emotional states: neutral (N), happiness (H), anger (A) and interrogative (I). The number of sentences is as follows: 34 for anger, 69 for happiness, 50 for interrogative and 99 for neutral states. The speakers have been amateur and have uttered each sentence several times from the template corpus. The emotional sentences with better quality have been selected from the recorded sentences.

The base features are 12 MFCCs, logarithm of energy and the velocity (Δ) and acceleration ($\Delta\Delta$) coefficients of them. The training corpus contains sentences of 14 speakers and the test corpus includes speech of 8 speakers. To study the effect of formant-related and pitch frequency-related features, they are added to the end of basic feature vector. Using three formant frequencies and pitch frequency, 16 supplementary features are calculated. These features contain the formants and pitch frequencies, the derivative and logarithm of them and their zero-mean values at each frame. To compute the zero-mean value, the mean value of that feature in each sentence is subtracted from the original value at each frame. These parameters and their abbreviations are listed in Table 1.

Data Classification Method Based on C-Support Vector Classification: The C-support vector classification (C-SVC) method is used in this section for pre-classification of the feature vectors of the four mentioned emotional states. In order to perform classification using the SVM method, three following steps are considered:

Step 1: The subject is first converted to $k(k\text{-}1)/2$ two-class problems. Each of the two-class problems is solved by C-SVC method which is described in the following. Thus, we have $k(k\text{-}1)/2$ separator lines for each two-class problem.

Step 2: Each test data is applied to $k(k\text{-}1)/2$ separators. Thus, each class competes in k-1 steps.

Step 3: The class which is the most successful in Step 2 is considered as the class of that test data.

As it has been mentioned in Step 1, the two-class problem is solved by C-SVC method:

In this method, suppose the training vectors as $x_i \in R^n$, i=1..., m in two classes and a vector $y \in R^l$ such that $y_i \in \{-1,1\}$. In this way, C-SVC solves the following problem [39]:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i \quad (1)$$

Subject to $y_i (w^T \phi(x_i)+b) \geq 1-\xi_i, \ \xi_i \geq 0, \quad i = 1,...,m.$

Its dual is

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \quad (2)$$

Subject to $y^T \alpha = 0, 0 \leq \alpha_i \leq C; \quad i = 1,....,m$

Where e is the vector of all ones, $C > 0$ is the upper bound, Q is an m by m positive semi-definite matrix.

Here, the training vectors, x_i, are mapped into a higher (maybe infinite) dimensional space by the function φ. The decision function is as follows:

$$\text{sgn} \ (\sum_{i=1}^{m} y_i \alpha_i K(x_i, x) + b) \quad (3)$$

Feature Ranking Methods: Fisher score, linear SVM and mutual information are the three feature ranking methods that are used in this paper. So, we can select the most important features among the 55 features introduced in Table 1. These methods are reviewed briefly as follow.

Fisher Score: F-score (Fisher score) is a simple and effective criterion to measure the discrimination between a feature and the label. Based on the statistical characteristics, it is independent of the classifiers. Following [40], a variant of F-score is used here. Given training instances x_i; i = 1,..., m, the F-score of the j^{th} feature is defined as follows:

$$F(j) \equiv \frac{(\bar{x}_j^{(+)} - \bar{x}_j)^2 + (\bar{x}_j^{(-)} - \bar{x}_j)^2}{\frac{1}{n_+ - 1} \sum_{i=1}^{n_+} (x_{i,j}^{(+)} - \bar{x}_j^{(+)})^2 + \frac{1}{n_- - 1} \sum_{i=1}^{n_-} (x_{i,j}^{(-)} - \bar{x}_j^{(-)})^2} \quad (4)$$

Where n_+ indicates the set of samples that are located in class +1 and n_- indicates the set of samples that are located in class -1, respectively. $\bar{x}_j$ is the j_{th} average feature, $\bar{x}_j^{(+)}$ indicates j_{th} average feature in class +1 and $\bar{x}_j^{(-)}$ indicates j_{th} average feature in class -1. $x_{i,j}^{(-)}$ indicates j_{th} average feature of the i_{th} positive sample.

Thus, the numerator indicates inter-class variance. The denominator is the sum of the variance within each class. So, if the numerator is increased or the denominator is decreased, then a larger F-score is achieved which means the feature is more discriminative. This means that if the two classes are more different, then a higher score for feature ranking is achieved.

Linear SVM: Support vector machines (SVMs) are useful for data classification. An SVM finds a separating hyperplane with the maximal margin between two classes of data. Given a set of instance-label pairs (x_i,y_i), $x_i \in R^n$, $y_i \in \{-1,1\}$, i = 1,...,m, SVM solves the following unconstrained optimization problem:

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \zeta(w, b; x_i, y_i) \quad (5)$$

Where ζ is a loss function and $C \geq 0$ is a penalty parameter on the training error. We use $\max(1-y_i(w^T\phi(x_i)+b),0)^2$ which is known as L2-loss SVM. ϕ is a function that maps training data into the higher dimensional spaces.

For any testing instance x, the decision function (predictor) is as follows:

$$f(x) = \text{sgn}(w^T \phi(x) + b) \quad (6)$$

A kernel function $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ may be used to train the SVM. In linear SVM, we have $K(x_i, x_j) = x_i^T x_j$. The LIBSVM tool, as a library for SVMs [41], is used in our simulations.

After obtaining a linear SVM model, $w \in R^n$ in (5) can be used to decide the relevance of each feature [42]. If $|w_j|$ has a large value, then the j^{th} feature plays an important role in the decision function (6). Only w in the linear SVM model has this indication, so this approach is restricted to linear SVM. We thus rank the features according to $|w_j|$. The procedure is as follows:

Input: Training sets, $(x_i; y_i)$; i = 1,...,m.
Output: Sorted feature ranking list.

- Use grid search to find the best parameter C.

- Train a L2-loss linear SVM model using the best C.
- Sort the features according to the absolute values of weights in the model.

Mutual Information: Mutual information is a measure of the dependence between random variables. It is always symmetric and non-negative. It is zero, if and only if the variables are statistically independent. The mutual information between class C and the discrete features $U=(u_1,u_2,...,u_d)$ can be calculated using (7). So, we can measure the mutual information between class C and the continuous feature space $X=(x_1,x_2,...,x_d)$ as follow:

$$I(X,C)=\sum_{C} p(c)\int_{-\infty}^{+\infty} p(X/c)\log\frac{p(X/c)}{P(X)}dX$$
$$=\sum_{C} p(c)\int_{-\infty}^{+\infty} p(X/c)\log p(X/c)dX-\int_{-\infty}^{+\infty} p(X)\log p(X)dX \quad (7)$$

If we want to calculate the mutual information between class C and the continuous feature space $X=(x_1,x_2,...,x_d)$, defined in (7), the main difficulty is to calculate the following term:

$$\Phi=-\int_{-\infty}^{\infty} f(X)\log f(X)dX \quad (8)$$

Φ is the entropy of $f(X)$ and plays an important role in the mutual information. We can not obtain the exact analytic solution of (8), especially if $f(X)$ has more than one component in a mixture model, but we can use the asymptotic formula to approximate (8). If sufficient data is existed, then Φ in (8) is approximated using the following equation:

$$\Phi\approx-\frac{1}{N}\sum_{n=1}^{N}\log f(X_n) \quad (9)$$

Where N is the size of data $\{X_n\}$.

In this section, the algorithm of optimum feature ranking using mutual information method is explained. This algorithm has four steps. First, suppose that a feature space $F^{(n)}$ contains n features ($n\geq 2$) and with initially set $m=n$. The steps of feature ranking algorithm are as follow [43]:

Step 1: Computing mutual information between the subset feature space and the class: Delete each feature $f_i(i=1,2,...,m)$ from the feature space $F^{(n)}$ to get $F_{-fi}^{(m)}$, where $F_{-fi}^{(m)}$ indicates the feature space $F^{(m)}$ with the feature f_i removed and calculate the mutual information between the feature space $F_{-fi}^{(m)}$, and the class using (7) and (9). Perform this step for each feature in the feature space $F^{(m)}$.

Step 2: Finding the least important feature in the feature space $F^{(m)}$. The feature f_i = max $\{I(F^{(m)},(C))\}$ is the least important one and the rank of feature f_j is m.

Step 3: Deleting feature f_i from the feature space $f_j^{(m)}$ and set $m=m-1$.

Step 4: If $m>1$ go to Step 1, otherwise stop and the last feature is ranked as the first.

After the feature ranking of all features, choose the following feature space which has the maximum mutual information with respect to the class:

$$F^{(m)}=\max_{m}\ \{I(F^{(m)},C)\} \quad (10)$$

Feature Ranking Results: In this section, the results of feature ranking using the three mentioned methods are reported. For this purpose, the investigations are performed for three sets of features: a) total features (F1-F55), b) 39 base features (F1-F39), c) 16 supplementary features (F40-F55). The results of ranking for the mentioned feature sets are reported in Tables 2 to 4. It is noted that the order of ranking is from left to right in consecutive rows.

As shown in Table 2, for the total of 55 features, F2, F3, F11, F13 and F46 are the most significant features that have been ranked by FS and MI-based methods. Also as can be seen in Table 3, for 39 base features, F2, F3, F11 and F13 are the most significant features that are ranked by FS, linear SVM and MI-based methods. The reported results in Table 4 show that for 16 supplementary features, F42, F46, F52 and F55 are the most significant features that are ranked by FS, linear SVM and MI-based methods.

Proposed Emotion Recognition System: The block diagram of the proposed emotion recognition system is shown in Fig. 1. As can be seen, based on the previous sections, C-support vector classification and three feature ranking methods are applied to the extracted features.

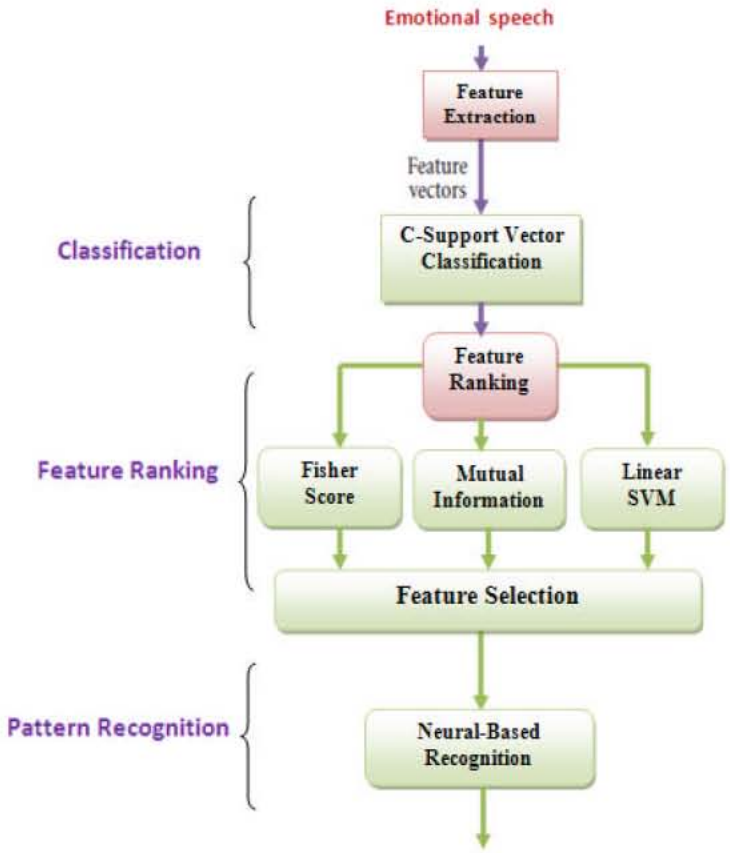

Fig. 1: Block diagram of proposed emotion recognition from speech

Table 1: List of 55 features used for emotion recognition from speech

Features	Abbreviations	Notation in feature set
12 MFCCs	$C_1, C_2, \ldots, C_{12}$	F1-F12
Logarithm of energy	LE	F13
12 first derivatives of MFCCs	$dC_1, dC_2, \ldots, dC_{12}$	F14-F25
First derivative of LE	dLE	F26
12 second derivatives of MFCCs	$ddC_1, ddC_2, \ldots, ddC_{12}$	F27-F38
Second derivative of LE	ddLE	F39
Pitch and formants	F_0, F_1, F_2, F_3	F52, F40-F42
First derivatives of F_0-F_3	dF_0, dF_1, dF_2, dF_3	F53, F43-F45
Logarithms of F_0-F_3	$logF_0, logF_1, logF_2, logF_3$	F54, F46-F48
Zero-mean values of F_0-F_3	ZF_0, ZF_1, ZF_2, ZF_3	F55, F49-F51

Table 2: 55 ranked features when using three feature ranking methods

Feature ranking method	Ranked features							
Fisher Score	F13	F10	F3	F2	F40	F46	F49	F11
	F55	F12	F41	F50	F45	F54	F52	F26
	F4	F15	F28	F48	F6	F44	F51	F47
	F30	F17	F9	F24	F37	F53	F1	F42
	F18	F39	F25	F31	F19	F20	F5	F23
	F21	F38	F35	F16	F36	F14	F29	F32
		F22	F34	F7	F43	F27	F8	F33

Table 2: 55: Continue

Feature ranking method	Ranked features							
Linear SVM	F5	F51	F13	F50	F52	F40	F41	F42
	F2	F45	F6	F3	F7	F12	F49	F4
	F47	F48	F43	F44	F8	F1	F9	F11
	F15	F17	F18	F53	F55	F54	F46	F10
	F22	F29	F30	F28	F25	F31	F16	F23
	F20	F35	F39	F34	F38	F14	F27	F26
		F24	F36	F33	F21	F19	F32	F37
Mutual Information	F11	F53	F2	F40	F46	F1	F3	F13
	F54	F49	F52	F47	F41	F48	F42	F6
	F21	F19	F39	F24	F9	F25	F4	F51
	F34	F55	F33	F18	F45	F15	F37	F20
	F43	F27	F16	F8	F10	F17	F23	F22
	F26	F28	F7	F5	F31	F50	F35	F12
		F38	F14	F44	F29	F32	F30	F36

Table 3: 39 ranked features when using three feature ranking methods

Feature ranking method	Ranked features							
Fisher Score	F10	F9	F6	F13	F2	F3	F7	F11
	F34	F1	F27	F39	F28	F8	F29	F4
	F25	F38	F33	F24	F19	F16	F30	F20
	F17	F32	F35	F5	F23	F15	F26	F12
		F22	F31	F14	F36	F18	F37	F21
Linear SVM	F1	F3	F11	F7	F2	F4	F13	F6
	F30	F28	F20	F9	F32	F5	F39	F19
	F26	F35	F17	F8	F10	F22	F37	F29
	F25	F34	F38	F23	F31	F33	F21	F18
		F12	F14	F27	F24	F16	F15	F36
Mutual Information	F2	F4	F19	F11	F25	F1	F3	F13
	F15	F37	F20	F21	F6	F39	F24	F9
	F10	F17	F23	F22	F34	F32	F33	F18
	F7	F5	F35	F31	F16	F27	F12	F8
		F14	F29	F38	F30	F36	F26	F28

Table 4: 16 ranked features when using three feature ranking methods

Feature ranking method	Ranked features							
Fisher Score	F42	F44	F49	F46	F52	F45	F55	F50
	F45	F54	F53	F47	F51	F43	F45	F43
Linear SVM	F46	F55	F47	F43	F42	F48	F52	F49
	F40	F50	F54	F53	F45	F41	F44	F51
Mutual Information	F48	F46	F50	F52	F51	F55	F41	F42
	F49	F45	F40	F54	F47	F44	F53	F43

Eight features are selected form the ranked features reported in Tables 2 to 4. For emotion recognition, two structures of ANNs (multi-layer perceptron (MLP) and radial basis function (RBF)) are used. For this purpose, a three-layer network has been used for modeling MLP. The activation functions of hidden and output layers have been selected as sigmoid and linear, respectively. Levenberg-Marquardt training function is used due to its fast convergence.

It is noted that RBF is a feedforward neural network with a radial symmetric Gaussian function as activation function of the hidden layer [44].

The number of input nodes of MLP and RBF ANNs are set to 8, due to the eight selected features from each of the three mentioned feature sets (Tables 2 to 4). The number of hidden nodes in two ANNs is set to 20. The number of output nodes is set to 4, due to the three emotional states and the interrogative state. The results of emotion recognition accuracy, when using different feature ranking methods by employing MLP and RBF, are shown in Tables 5 and 6, respectively.

The overall performance of the mentioned ANNs in emotion recognition using three feature ranking methods is also reported in Table 7.

Table 5: Performance of emotion recognition system when using MLP recognizer

	Accuracy using 8 selected features (%)					
	Among features F1-F39			Among features F40-F55		
	Feature ranking method			Feature ranking method		
State	MI	FS	L-SVM	MI	FS	L-SVM
Happiness	78.49	68.39	67.22	73.69	69.38	65.91
Neutral	85.21	72.77	73.55	79.17	79.34	61.67
Angry	87.29	72.11	68.13	69.38	67.54	66.65
Interrogative	71.58	83.92	82.28	64.95	60.27	62.96

Table 6: Performance of emotion recognition system when using RBF recognizer

	Accuracy using 8 selected features (%)					
	Among features F1-F39			Among features F40-F55		
	Feature ranking method			Feature ranking method		
State	MI	FS	L-SVM	MI	FS	L-SVM
Happiness	78.28	69.21	65.23	73.23	65.88	65.34
Neutral	84.77	71.98	73.26	78.55	79.11	61.56
Angry	86.31	73.67	67.51	69.21	66.47	66.83
Interrogative	71.32	82.31	81.25	64.72	60.25	62.98

Table 7: Overall emotion recognition performance of the proposed system

	Accuracy using 8 selected features (%)			
	Among features F1-F39		Among features F40-F55	
Feature ranking method	MLP recognizer	RBF recognizer	MLP recognizer	RBF recognizer
MI	80.64	80.17	71.80	71.43
FS	74.30	74.29	69.24	67.93
L-SVM	72.80	71.81	64.30	64.18

Table 8: Performance comparison of the proposed system with some recent researches

Emotional states	Family of features	Classifier(s)	Feature selection method(s)	Recognition rate (%)
Happiness, anger, sadness, neutral [30]	Pitch and its slope, formants, MFCCs	SVM, ANN	-	71, 42
Happiness, anger, tiredness, sadness, neutral [10]	Pitch, log energy, formants, MFCCs and their Δ and ΔΔ	GSVM[a]	-	41
Happiness, anger, sadness, neutral [12]	Pitch, sub-band energies, MFCCs, LPC[b]	Multi-class SVM	-	80
Happiness, anger, sadness, fear, neutral [45]	Pitch, intensity, zero crossing rate, spectral features	KNN	-	66
Anger, happiness, neutral, sadness, surprise [46]	Formants, pitch, energy, spectral features	MLB	SFFS	53.7(DES Database) 57.2(SUSASDatabase)
Anger, happiness, sadness, boredom, neutral [23]	LPC, MFCCs	KNN	FFS, BFS	79.6[c]
Anger, disgust, fear, happiness, neutral, sadness, surprise [32]	Pitch, energy, duration, MFCCs	MLB	PCA, LDA	53[c]
Happiness, anger, sadness, fear, neutral [24]	Pitch, speaking rate, formants, bandwidth	KNN	Instance-base learning	70
Happiness, anger, neutral and interrogative (proposed model)	MFCCs, log energy and their Δ and ΔΔ	MLP, RBF	MI	80.6, 80.2
Happiness, anger, neutral and interrogative (proposed model)	MFCCs, log energy and their Δ and ΔΔ	MLP, RBF	FS	74.3, 74.3
Happiness, anger, neutral and interrogative (proposed model)	MFCCs, log energy and their Δ and ΔΔ	MLP, RBF	L-SVM	72.8, 71.8
Happiness, anger, neutral and interrogative (proposed model)	Formant-related and pitch-related features	MLP, RBF	MI	71.8, 71.4
Happiness, anger, neutral and interrogative (proposed model)	Formant-related and pitch-related features	MLP, RBF	FS	69.2, 67.9
Happiness, anger, neutral and interrogative (proposed model)	Formant-related and pitch-related features	MLP, RBF	L-SVM	64.3, 64.2

[a] Gaussian SVM

[b] Linear Predictive Coding

[c] Maximum Emotion Recognition Rate

As can be seen, MI-based feature ranking method performs better as compared to two other feature ranking methods. Also, MLP neural recognizer performs slightly better as compared to RBF neural recognizer. The performance of the proposed system is also compared with some recent emotion recognition systems and reported in Table 8.

CONCLUSION

In this paper, three feature ranking methods have been used in the speech emotion recognition application. These methods were FS, L-SVM and MI. For this purpose, a rich feature set with the size of 55 has been used. Then two distinct feature subsets with the size of 39 and 16 have been selected from the mentioned feature set. To investigate the performance of system with a small-size input feature set, eight high-ranked features have been selected from each of the mentioned feature sets (with the size of 55, 39 and 16) and two types of neural networks (MLP and RBF) have been used for emotion recognition. Experimental results have shown that using MI-based feature ranking method and MLP recognizer results in emotion recognition rates above 80% by employing a small-size feature set.

REFERENCES

1. Yacoub, S., S. Simske, X. Lin and J. Burns, 2003. Recognition of Emotions in Interactive Voice Response Systems. In the Proceedings of the European Conference on Speech Communication and Technology, pp: 729-732.
2. Gharavian, D., M. Sheikhan and M. Janipour, 2010. Pitch in Emotional Speech and Emotional Speech Recognition Using Pitch Frequency. Majlesi Journal of Electrical Engineering, 4(1): 19-24.
3. Gharavian, D. and M. Sheikhan, 2010. Emotion Recognition and Emotion Spotting Improvement Using Formant-Related Features. Majlesi Journal of Electrical Engineering, 4(4): 1-8.
4. Gharavian, D., M. Sheikhan, A.R. Nazerieh and S. Garoucy, 2011. Speech Emotion Recognition Using FCBF Feature Selection Method and GA-Optimized Fuzzy ARTMAP Neural Network. Neural Computing and Applications, (Article in Press, DOI 10.1007/s00521-011-0643-1).
5. Sheikhan, M., M. Tebyani and M. Lotfizad, 1997. Continuous Speech Recognition and Syntactic Processing in Iranian Farsi Language. International Journal of Speech Technology, 1: 135-141.
6. Sheikhan, M., D. Gharavian and F. Ashoftedel, 2011. Using DTW Neural-Based MFCC Warping to Improve Emotional Speech Recognition. Neural Computing and Applications, (Article in Press, DOI 10.1007/s00521-011-0620-8).
7. Gharavian, D. and S.M. Ahadi, 2005. The Effect of Emotion on Farsi Speech Parameters: A Statistical Evaluation. In the Proceedings of the International Conference on Speech and Computer, pp: 463-466.
8. Gharavian, D. and S.M. Ahadi, 2006. Recognition of Emotional Speech and Speech Emotion in Farsi. In the Proceedings of International Symposium on Chinese Spoken Language Processing, 2: 299-308.
9. Gharavian, D., 2004. Prosody in Farsi Language and Its Use in Recognition of Intonation and Speech. Ph.D. Dissertation, Electrical Engineering Department, Amirkabir University of Technology, Tehran (in Farsi).
10. Kwon, O.W., K. Chan, J. Hao and T.W. Lee, 2003. Emotion Recognition by Speech Signals. In the Proceedings of the European Conference on Speech Communication and Technology, pp: 125-128.
11. Altun, H. and G. Polat, 2007. New Frameworks to Boost Feature Selection Algorithms in Emotion Detection for Improved Human-Computer Interaction. Brain Vision and Artificial Intelligent. Lecture Notes in Computer Science, 4729: 533-541.
12. Altun, H. and G. Polat, 2009. Boosting Selection of Speech Related Features to Improve Performance of Multi-Class SVMs in Emotion Detection. Expert Systems with Applications, 36: 8197-8203.
13. Dellaert, F., T. Polzin and A. Waibel, 1996. Recognizing Emotion in Speech. In the Proceedings of the International Conference on Spoken Language Processing, vol. 3, pp: 1970-1973.
14. Neiberg, D., K. Elenius and K. Laskowski, 2006. Emotion Recognition in Spontaneous Speech Using GMMs. In the Proceedings of the International Conference on Spoken Language Processing, pp: 809-812.
15. Ververidis, D. and C. Kotropoulos, 2006. Emotional Speech Recognition: Resources, Features and Methods. Speech Communication, 48: 1162-1181.
16. Nicholson, J., K. Takahashi and R. Nakatsu, 1999. Emotion Recognition in Speech Using Neural Networks. In the Proceedings of the International Conference on Neural Information Processing, 2: 495-501.

17. Park, C.H., D.W. Lee and K.B. Sim, 2002. Emotion Recognition of Speech Based on RNN. In the Proceedings of the International Conference on Machine Learning and Cybernetics, 4: 2210-2213.
18. Park, C.H. and K.B. Sim, 2003. Emotion Recognition and Acoustic Analysis from Speech Signal. In the Proceedings of the International Joint Conference on Neural Networks, vol. 4, pp: 2594-2598.
19. Shami, M. and W. Verhelst, 2007. An Evaluation of the Robustness of Existing Supervised Machine Learning Approaches to the Classification of Emotions in Speech. Speech Communication, 49: 201-212.
20. Morrison, D., R. Wang, L.C. de Silva, 2007. Ensemble Methods for Spoken Emotion Recognition in Call-Centers. Speech Communication, 49: 98-112.
21. Schuller, B., G. Rigoll and M. Lang, 2004. Speech Emotion Recognition Combining Acoustic Features and Linguistic Information in a Hybrid Support Vector Machine-Belief Network Architecture. In the proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp: 577-580.
22. Kao, Y. and L. Lee, 2006. Feature Analysis for Emotion Recognition from Mandarin Speech Considering the Special Characteristics of Chinese Language. In the Proceedings of the International Conference on Spoken Language Processing, pp: 1814-1817.
23. Pao, T., Y. Chen, J. Yeh and Y. Chang, 2008. Emotion Recognition and Evaluation of Mandarin Speech Using Weighted D-KNN Classification. International Journal of Innovative Computing, Information and Control, 4: 1695-1709.
24. Petrushin, V.A., 2000. Emotion Recognition in Speech Signal: Experimental Study, Development and Application. In the Proceedings of the International Conference on Spoken Language Processing, pp: 222-225.
25. Luengo, I., E. Navas, I. Hernaez and J. Sanchez, 2005. Automatic Emotion Recognition Using Prosodic Parameters. In the Proceedings of the European Conference on Speech Communication and Technology, pp: 493-496.
26. Schuller, B., G. Rigoll and M. Lang, 2003. Hidden Markov Model-Based Speech Emotion Recognition. In the Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 2: 1-4.
27. Bosch, L., 2003. Emotions, Speech and the ASR Framework. Speech Communication, 40: 213-225.
28. Lee, C., S. Narayanan and R. Pieraccini, 2002. Classifying Emotions in Human-Machine Spoken Dialogs. In the Proceedings of the International Conference on Multimedia and Expo, 1: 737-740.
29. Petrushin, V., 1999. Emotion in Speech: Recognition and Application to Call Centers. In the Proceedings of Artificial Neural Networks in Engineering Conference, pp: 7-10.
30. Yu, F., E. Chang, Y.Q. Xu and H.Y. Shum, 2001. Emotion Detection from Speech to Enrich Multimedia Content. In the Proceedings of the Second IEEE Pacific-Rim Conference on Multimedia, pp: 550-557.
31. Sidorova, J., 2009. Speech Emotion Recognition with TGI+.2 Classifier. In the Proceedings of the EACL, Student Research Workshop, pp: 54–60.
32. Haq, S., P.J.B. Jackson and J. Edge, 2008. Audio-Visual Feature Selection and Reduction for Emotion Classification. In the Proceedings of the International Conference on Auditory-Visual Speech Processing, pp: 185-190.
33. Sheikhan, M., V. Tabataba Vakili and S. Garoucy, 2009. Complexity Reduction of LD-CELP Speech Coding in Prediction of Gain Using Neural Networks. World Applied Sciences Journal, 7 (Special Issue of Computer & IT): 38-44.
34. Sheikhan, M., V. Tabataba Vakili and S. Garoucy, 2009. Codebook Search in LD-CELP Speech Coding Algorithmin Based on Multi-SOM Structure. World Applied Sciences Journal, 7(Special Issue of Computer & IT): 59-68.
35. Sheikhan, M. and S. Garoucy, 2010. Reducing the Codebook Search Time in G.728 Speech Coder Using Fuzzy ARTMAP Neural Networks. World Applied Sciences Journal, 8: 1260-1266.
36. Sheikhan, M. and S. Garoucy, 2010. Hybrid VQ and Neural Models for ISF Quantization in Wideband Speech Coding. World Applied Sciences Journal, 10(Special Issue of Computer & Electrical Engineering): 59-66.
37. Sheikhan, M., D. Gharavian and A. Eslamzadeh, 2010. Enhancement of LPC-10 Speech Coder Using LSP Parameters and Neural Vector Quantizers. World Applied Sciences Journal, 10 (Special Issue of Computer & Electrical Engineering): 41-48.
38. Sheikhan, M. and S. Garoucy, 2011. Prediction of Gain in LD-CELP Using Hybrid Genetic/PSO-Neural Models. Journal of Advanced Researches in Computer, 1(3): 1-12.

39. Mathur, A. and G.M. Foody, 2008. Multiclass and Binary SVM Classification: Implications for Training and Classification Users. IEEE Geoscience and Remote Sensing Letters, 5: 241-245.
40. Chen, Y.W. and C.J. Lin, 2006. Combining SVMs with Various Feature Selection Strategies. In I. Guyon, S. Gunn, M. Nikravesh and L. Zadeh, Editors, Feature Extraction, Foundations and Applications, Springer.
41. Chang, C.C. and C.J. Lin, 2010. A Library for Support Vector Machines. Initial Version: 2001, Last Updated: March 6, 2010.
42. Guyon, I., J. Weston, S. Barnhill and V. Vapnik, 2002. Gene Selection for Cancer Classification Using Support Vector Machines. Machine Learning, 46: 389-422.
43. Cang, S. and D. Partridge, 2004. Feature Ranking and Best Feature Subset Using Mutual Information. Neural Computing and Applications, 13: 175-184.
44. Suresh, S., N. Sundararajan and P. Saratchandran, 2008. A Sequential Multi-Category Classifier Using Radial Basis Function Networks. Advances in Neural Information Processing Systems, 1: 1345–1358.
45. Rong, J., G. Li and Y.P. Chen, 2009. Acoustic Feature Selection for Automatic Emotion Recognition from Speech. Information Processing and Management, 45: 315-328.
46. Ververidis, D. and C. Kotropoulos, 2006. Fast Sequential Floating Forward Selection Applied to Emotional Speech Features Estimated on DES and SUSAS Data Collections. In the Proceedings of the European Signal Processing Conference, pp: 1-5.