Research Article

Reordering Features with Weights Fusion in Multiclass and Multiple-Kernel Speech Emotion Recognition

Xiaoqing Jiang,[1,2] **Kewen Xia,**[1] **Lingyin Wang,**[2] **and Yongliang Lin**[1,3]

[1]*School of Electronics and Information Engineering, Hebei University of Technology, Tianjin 300401, China*
[2]*School of Information Science and Engineering, University of Jinan, Shandong, Jinan 250022, China*
[3]*Information Construction and Management Center, Tianjin Chengjian University, Tianjin 300384, China*

Correspondence should be addressed to Kewen Xia; kwxia@hebut.edu.cn

The selection of feature subset is a crucial aspect in speech emotion recognition problem. In this paper, a Reordering Features with Weights Fusion (RFWF) algorithm is proposed for selecting more effective and compact feature subset. The RFWF algorithm fuses the weights reflecting the relevance, complementarity, and redundancy between features and classes comprehensively and implements the reordering of features to construct feature subset with excellent emotional recognizability. A binary-tree structured multiple-kernel SVM classifier is adopted in emotion recognition. And different feature subsets are selected in different nodes of the classifier. The highest recognition accuracy of the five emotions in Berlin database is 90.549% with only 15 features selected by RFWF. The experimental results show the effectiveness of RFWF in building feature subset and the utilization of different feature subsets for specified emotions can improve the overall recognition performance.

1. Introduction

Feature selection is a crucial aspect in pattern recognition problems. In multiclass SVM classifier, for example, the structure of the classifier can be one-to-one, one-to-all, hierarchy, or tree structure, so several SVM nodes or models exist in the multiclass classifier [1–3]. There are two questions in speech emotion recognition (SER): (1) how to seek the optimal feature subset from the acoustic features; (2) whether the same acoustic feature subset is proper in all nodes of the multiclass classifier. These questions are researched in this paper. A novel algorithm named Reordering Features with Weights Fusion (RFWF) is proposed to select feature subsets. And in emotion recognition procedure, different feature subsets are adopted in SVM nodes to recognize different emotions.

In SER field, the dimension of feature set ranges from tens to hundreds. However, the increasing dimension does not mean a radical improvement of the recognition accuracy, because the variety and redundancy between more and more features influence the overall performance and complexity

of the system [4]. And there is not a categorical assertion about the most effective feature set in SER nowadays. Feature selection algorithms used in machine learning widely can choose the optimal feature subset with the least generalization error. There are three types of feature selection methods: the wrapper method, the embedded method, and the filter method [5]. Compared with the wrapper method and the embedded method, the filter method is simpler and faster in calculation and its learning strategy is more robust to overfitting. Additionally, because the selection result of the filter method is independent of the learning model, the filter method can be adopted in a variety of leaning tasks. The criteria in filter methods mainly focus on the relevance, the redundancy, and the complementarity. For example, Joint Mutual Information (JMI) [6] considers the relevance between features and classes. Fast correlation-based filter (FCBF) [7] takes the redundancy between features into account. Max-Relevance Min-Redundancy (MRMR) [8] gives consideration to both relevance and redundancy to find the balance between the two properties. In Conditional Information Feature Extraction (CIFE) [9], the information

TABLE 1: Feature set.

Type	Feature	Statistic parameters
Prosodic feature	Pitch	Maximum (1), minimum (2), range (3), mean (4), Std (5), first quartile (6), median (7), third quartile (8), interquartile range (9)
	Energy	Maximum (10), minimum (11), range (12), mean (13), Std (14), first quartile (15), median (16), third quartile (17), interquartile range (18)
	Time	Total frames (19), voiced frames (20), unvoiced frames (21), ratio of voiced frames versus unvoiced frames (22), ratio of voiced frames versus total frames (23), ratio of unvoiced frames versus total frames (24)
Voice quality feature	Formant	F_1:mean (25), Std (26), median (27) F_2: mean (28), Std (29), median (30) F_3: mean (31), Std (32), median (33)
Spectral feature	MFCC	12 MFCC (34–45)

provided by the features is divided into two parts: the class-relevant information that benefits the classification and the class-redundant information that disturbs the classification. And the key idea of Double Input Symmetrical Relevance (DISR) [10] is the utilization of symmetric relevance to consider the complementarity between two input features. In the SER field, various feature selection criteria are often adopted [7, 11–13], and different criterion emphasizes different aspects. Reordering Features with Weights Fusion (RFWF) algorithm proposed in this paper aims to consider relevance, redundancy, and complementarity comprehensively.

Traditionally the same feature subset is adopted in all emotional classes for training and testing [14]. In [11], different feature subsets are adopted on two emotional speech databases, but the emotional recognizability of the features to different emotions has not been considered. Research has shown that acoustic features have different recognizability to specific emotions. For example, pitch related features are usually essential to classify happy and sad [15], while they are often weak in the recognition between happy and surprise because of their high values in these emotions [16]. In order to improve the performance of the whole system, different feature subsets are selected and adopted on the different nodes of the multiclass classifier in this paper.

The content of the paper is arranged as follows: Section 2 gives the basic concepts of filter feature selection and the method of RFWF; Section 3 introduces the structure of the multiclass and multiple-kernel SVM classifier; Section 4 is the analysis of experiments including the results of RFWF and recognition accuracies of emotions; and the final section is devoted to the conclusions.

2. Features and Feature Selection Methods

2.1. Acoustic Features. Speech acoustic features usually used in SER are the prosodic features, voice quality features, and spectral features. In this paper, 409 utterances in Berlin database [17] of 5 emotions including happy (71 samples), angry (127 samples), fear (69 samples), sad (63 samples), and neutral (79 samples) are studied. These samples are separated into training and testing categories randomly. The training samples are 207 including happy (36 samples), angry (64

samples), fear (35 samples), sad (32 samples), and neutral (40 samples), and the rest 202 ones are the test samples.

Pitch, energy, time, formant, and Mel Frequency Cepstrum Coefficient (MFCC) and their statistics parameters are extracted. The total dimension the feature set is 45. Table 1 lists the acoustic features and their sequence indices in this paper.

2.2. Mathematical Description of Feature Selection. Relevance, redundancy, and complementarity are considered in feature selection methods. If a feature can provide information about the class, the relevance exists between the feature and class. The redundancy is based on the dependency between the selected and unselected features. And complementarity means that the interaction between an individual feature and the selected feature subset is beneficial to the classification. Complementarity is important in the cases of null relevance, such as XOR problem [10, 18].

The concepts of information theory, such as mutual information denoted by I and entropy denoted by H, are widely used in feature selection. Mathematically, F_i ($i = 1, \ldots, 409$) is the feature vector of ith sample, and the $f_{i,j}$ is the jth feature of the ith sample in the feature set $\mathbf{F}$. The selected subset and unselected subset are $\mathbf{F}_s$ and $\mathbf{F}_{-s}$ with the mathematic relation of $\mathbf{F}_s \cap \mathbf{F}_{-s} = \varnothing$ and $\mathbf{F}_s \cup \mathbf{F}_{-s} = \mathbf{F}$. C_n, $n = 1, \ldots, 5$ is the specified emotion in Berlin database. In the following content, the mathematical description of relevance, redundancy, and complementarity is interpreted through the introduction of MRMR and DISR.

In MRMR $\mathbf{F}_s = \{f_p\}$, $p = 1, \ldots, d$ and $f_q \in \mathbf{F}_{-s}, q = 1, \ldots, 45 - d$, the relevance term $u_q = I(f_q; C)$ and the redundancy term $z_q = 1/d \sum_{f_p \in \mathbf{F}_s} I(f_p; f_q)$ are used in the criterion:

$$f_q^{\text{MRMR}} = \arg\max_{f_q \in \mathbf{F}_{-s}} \left\{ u_q - z_q \right\}, \tag{1}$$

where $I(f_q; C)$ can represent the relevance between an unselected feature and the class and $I(f_p; f_q)$ can represent the redundancy between the unselected and selected features. The detailed computation can be found in [8].

The key idea of DISR depends on the consideration of the second average sub-subset information criterion in (2) to

consider the complementarity between an unselected feature f_q and a selected feature f_p given the specific class C.

$$f_q^{\text{DISR}} = \arg\max_{f_q \in \mathbf{F}_{-s}} \left\{ \sum_{f_p \in \mathbf{F}_s} I\left(f_{p,q}; C\right) \right\}. \tag{2}$$

Equation (2) also can be modified by a normalized relevance measure named symmetric relevance calculated by the following:

$$\text{SR} = \frac{I\left(f_{p,q}; C\right)}{H\left(f_{p,q}; C\right)}. \tag{3}$$

In DISR, $I(f_{p,q}; C)$ is the complementarity calculated by

$$I\left(f_{p,q}; C\right) = I\left(f_p; C\right) + I\left(f_q; C\right) - A\left(f_p; f_q; C\right), \tag{4}$$

where $A(f_p; f_q; C)$ stands for the interaction among f_p, f_q, and C. From its general meaning, for n sets of random variables $X_1, X_2, \ldots, X_n$, the interaction can be defined as

$$A\left(X_1, X_2, \ldots, X_n\right) = \sum_{k=1}^{n} \sum_{S \subseteq \{1,\ldots,n\}: S=k} (-1)^{k+1} H\left(X_S\right). \tag{5}$$

The detailed definition and proof can be found in [10].

2.3. Reordering Features with Weights Fusion. For the comprehensive consideration of relevance, redundancy, and complementarity, the following criterion named Reordering Features with Weights Fusion (RFWF) is proposed to fuse the intrinsic properties of the features:

$$f_q^{\text{RFWF}} = \arg\max_{f_q \in \mathbf{F}_{-s}} \left\{ W_1\left(I\left(f_q; C\right)\right) \right.$$
$$\left. + W_2\left(\sum_{f_p \in \mathbf{F}_s} I\left(f_{p,q}; C\right)\right) - W_3\left(\sum_{f_p \in \mathbf{F}_s} I\left(f_p; f_q\right)\right) \right\}, \tag{6}$$

where W_1, W_2, and W_3 are the fusing weights of the unselected feature f_q and they are combined in (6) to reflect the contribution of f_q to given class. The procedure of RFWF algorithm described in the following is illustrated in Figure 1:

(1) $L_m(f_q)$ $(m = 1, 2, 3)$ is the sequence number of the feature f_q ranked in order of the values of $I(f_q; C)$, $\sum_{f_p \in \mathbf{F}_s} I(f_{p,q}; C)$ and $\sum_{f_p \in \mathbf{F}_s} I(f_p; f_q)$, respectively. If the dimension of the feature set is 45, $L_m(f_q)$ is an integer value ranging within 1~45. For example, if the $I(f_q; C)$ is the largest, $L_1(f_q)$ is 1. And if the $\sum_{f_p \in \mathbf{F}_s} I(f_{p,q}; C)$ is the lowest, $L_2(f_q)$ is 45. The initial selected feature f_p in $\mathbf{F}_s$ is confirmed by the largest value of $I(f_p; C)$.

(2) Weighted values can be calculated by the following formula:

$$W_m\left(f_q\right) = 45 - L_m\left(f_q\right) + 1,$$
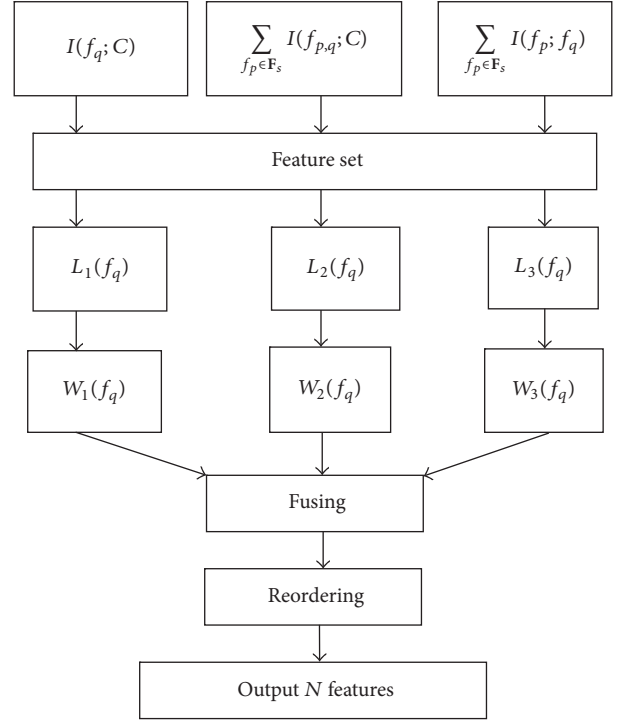$$m = 1, 2, 3, \quad q = 1, \ldots, 45. \tag{7}$$

FIGURE 1: Procedure of RFWF algorithm.

For example, if $L_1(f_q)$ is 1, the corresponding weight about the relevance between feature and class is $W_1(I(f_q; C)) = 45$.

(3) All of the features can be reordered by the fusing result using W_1, W_2, and W_3.

(4) The top N features can be selected to construct the optimal feature subset.

Because this algorithm fuses three weights to consider the contribution of features in the classification and a reordering process exists in the process, the algorithm is named Reordering Features with Weights Fusion (RFWF).

3. Multiclass and Multiple-Kernel SVM Classifier with Binary-Tree Structure

The Support Vector Machine (SVM) is a discriminative classifier proposed for binary classification problem and based on the theory of structural risk minimization. The performance of a single-kernel method depends heavily on the choice of the kernel. If a dataset has varying distributions, a single kernel may not be adequate. Kernel fusion has been proposed to deal with this problem [19].

The simplest kernel fusion is a weighted combination of M kernels:

$$\mathbf{K} = \sum_{s=1}^{M} \mu_s \mathbf{K}_s, \tag{8}$$

where μ_s is the optimal weights and $\mathbf{K}_s$ is the sth kernel matrix. The selection of μ_s is an optimal question, and the
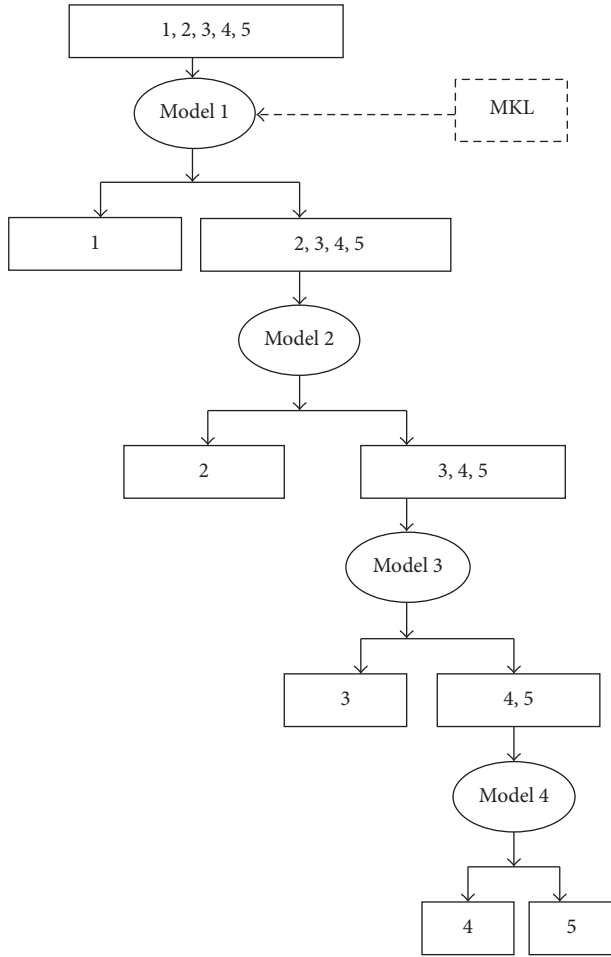
Figure 2: The binary-tree structured multiclass and multiple-kernel SVM classifier.

objective function and constraints of the problem can be formulated as the Semidefinite Programming (SDP) form. The detailed proof of can be found in [20].

In this paper, a multiple-kernel SVM classifier with an unbalance binary-tree structure illustrated in Figure 2 is adopted. In Figure 2, there are five emotions to be recognized. The first classifying node (Model 1) is improved by multiple-kernel SVM to recognize the most confusable emotion while the subsequent classifying nodes still retain single-kernel SVM. This arrangement attributes the reduction of recognition error accumulation and the computing complexity required in the calculation of the multiple-kernel matrices for all nodes.

According to the previous works [2, 11, 21, 22], happy is the most confusable emotion and its recognition accuracy is the main factor influencing the total performance in Berlin database. Thus in the classifier shown in Figure 2, happy is 1, angry is 2, fear is 3, neutral is 4, and sad is 5. Feature subset selected by RFWF is adopted in the SVM training and testing, where Model 1 is learned by multiple kernels. Models 2, 3, and 4 are still single-kernel SVM models.

4. Experiments and Analysis

4.1. The Experimental Results of RFWF. Table 2 lists the reordering results of features for the four SVM models in Figure 2 according to the fusing results of W_1, W_2, and W_3. In Table 2, the numbers are the indices of the features listed in Table 1.

It is clear that, in the four SVM models, the contribution of different features to emotional recognizability is distinct. For example, the standard deviation of pitch (feature sequence index is 5) is the most essential feature to classify happy and the other emotions in Berlin database, while ratio of voiced frames versus total frames (feature sequence index is 23) is the most important feature to recognize neutral and sad. The results show that it is necessary to adopt different feature subsets to recognize different emotions.

4.2. Experimental Results of SER and Analysis. In the SER experiments, LibSVM package is adopted. Three basis Radial Basis Functions (RBF) kernel functions with parameters of $\gamma_1 = 0.01$, $\gamma_2 = 0.1$, $\gamma_3 = 1$ are combined in Model 1. YALMIP toolbox is used to solve the SDP problem and find three μ_s with the features listed in Table 1. In single-kernel SVM models, the value of γ is $1/k$, where k is the selected feature number in the recognition procedure. When the selected feature number is specified, the same γ is adopted for all single-kernel models.

Recognition accuracies, Root Mean Square Error (RMSE) and Maximum Error (MaxE) are used to evaluate the performance of the SVM classifier. RMSE and MaxE are calculated by following equations:

$$P_{\text{RMSE}} = \sqrt{\frac{1}{5}\sum_{i=1}^{5} e_i{}^2}$$

$$Q_{\text{MaxE}} = \max\{e_i\},$$

(9)

where e_i is the recognition error (%) of ith emotion. Obviously, the higher the recognition accuracies and the lower the values of RMSE and MaxE, the better the performance of the classifier.

If the dimension of feature subset is N, then the top N features in the Table 2 are selected to construct the feature subset. N ranging within 1–45 achieves different recognition performance. Figure 3 plots the curves of total emotion recognition accuracies of MRMR, DISR, and RFWF features selection algorithms, respectively, where RFWF is adopted in multiple-kernel (RFWF MK) and single-kernel (RFWF SK) SVM classifiers. Different feature subsets are selected for Models 1–4. Figure 4 gives the RMSE and MaxE corresponding to Figure 3. In Figures 3 and 4, the horizontal axis describes the different number of the selected features or the dimension of the selected feature subset. Table 3 lists the detailed experimental data of the highest total accuracies of MRMR, DISR, RFWF MK, and RFWF SK methods.

The recognition results show DISR and MRMR algorithms reach their highest accuracies with 39 features. However, the highest accuracy of RFWF MK is 90.594% with only 15 features. The accuracies of DISR and MRMR are 70.792%

TABLE 2: RFWF results of the feature set in different models of the classifier.

	Indices of features														
	5	37	38	42	8	1	9	20	40	33	26	3	4	45	18
Model 1	44	15	24	21	13	29	10	6	7	14	16	19	23	17	35
	41	22	34	25	2	12	31	43	36	39	28	30	32	11	27
	36	34	20	43	42	7	5	14	18	41	16	8	4	15	1
Model 2	6	33	37	24	19	23	45	9	3	31	13	17	21	10	32
	22	39	40	2	30	28	35	12	38	11	29	27	44	26	25
	4	20	43	35	36	45	8	16	7	18	9	23	34	11	25
Model 3	19	21	2	6	5	22	44	27	10	24	38	32	29	42	12
	26	40	14	17	39	13	31	15	37	41	28	30	1	33	3
	23	7	39	41	14	10	15	19	24	22	13	21	40	35	11
Model 4	16	42	25	27	38	8	18	33	36	43	12	20	6	29	5
	30	17	31	4	2	37	28	9	44	1	26	34	3	32	45

TABLE 3: Emotion recognition accuracies of different feature selection methods with the best feature number.

Selection methods	Feature number	SER accuracies (%)					
		Total	Angry	Fear	Happy	Neutral	Sad
DISR	39	88.614	93.651	79.412	97.143	74.359	96.774
MRMR	39	90.099	92.063	67.647	97.143	94.872	96.774
RFWF MK	15	90.594	90.476	79.412	97.143	87.179	100
RFWF SK	39	79.208	95.238	73.529	28.571	87.179	100

and 77.723%, respectively, when the selected feature number is 15. When the selected feature number is 45, it means that no feature selection algorithms are utilized. In this situation, the performances of DISR, MRMR, and RFWF MK are the same and the total accuracy is 83.663%. These results show that the RFWF algorithm has the best performance with the lowest dimension of feature subset. The corresponding RMSE and MaxE curves of RFWF are the lowest when the selected feature number N is below 30. If the dimension of the feature subset increases, the three feature selection methods with multiple-kernel classifier have similar performance. This is mainly because RFWF uses the same weighing methods for the relevance, redundancy, and complementarity. From this aspect, it is an average strategy in the procedure of weights fusion. The results show that when the dimension feature subset is close to 45, the RFWF degrades to handle with the complex inherent properties between the features and more optimal feature fusion method should be studied.

The highest total accuracy of RFWF SK is 79.208%, which is much lower than accuracy of RFWF MK. The recognition accuracies of happy are 97.143% steadily in the three methods when Model 1 is improved by multiple-kernel SVM. The experimental results demonstrate that multiple-kernel classifier can solve the confusion between happy and other emotions effectively, which cannot be dealt with by single-kernel SVM. The highest SER accuracies of RFWF MK can be compared to the results of the Enhanced Sparse Representation Classifier (Enhanced-SRC) in [11] and feature fusion based on MKL in [23]. The experimental comparison is listed in Table 4, where the symbol "N" denotes no relating experimental results in the reference.
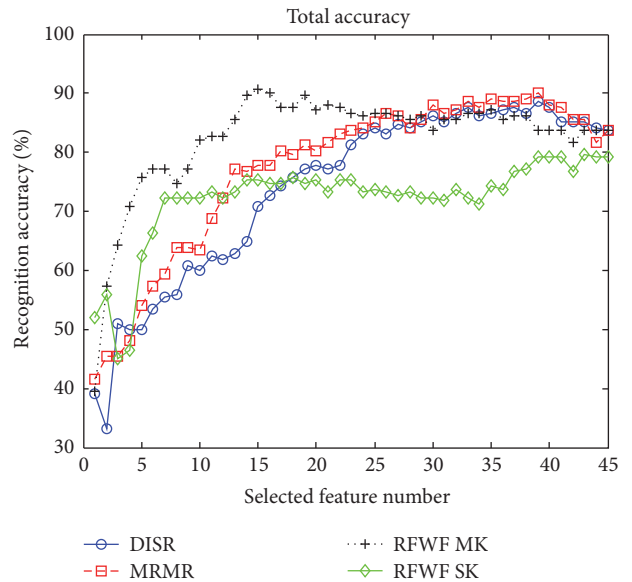

FIGURE 3: Emotional recognition accuracies.

If Models 2–4 use the same feature subset as in Model 1 with the dimension of 15, the accuracy of RFWF MK is only 63.861%. And when all models use the same feature subset of 39, the highest accuracy of RFWF MK is 85.149%. The data confirms that the utilization of the same feature subset in all models influences the emotion recognition performance negatively. These experimental results demonstrate that different feature subset is necessary in the recognition of

TABLE 4: Comparison on SER accuracies (%) of 5 emotions in Berlin database.

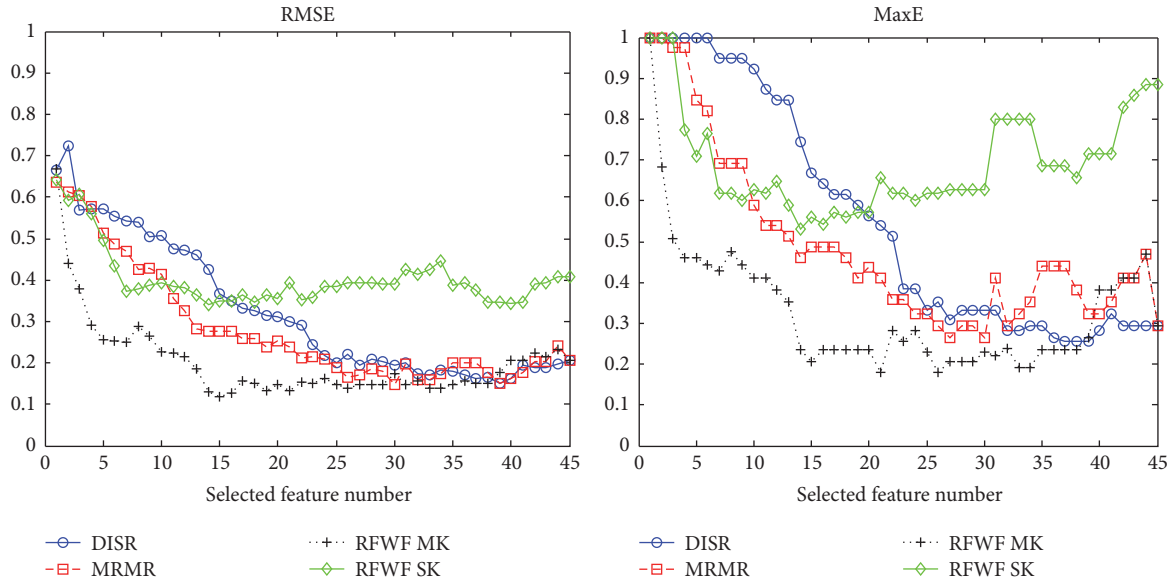| Methods	Anger	Fear	Happy	Neutral	Sad
Enhanced-SRC	98.55	83.16	57.73	70.08	96.71
Feature fusion based on MKL	81	83	N	65	95
RFWF MK	90.476	79.412	97.143	87.179	100

FIGURE 4: RMSE and MaxE.

different emotions, which also indicates the difficulty to build a robust and effective feature subset for all emotions.

5. Conclusions

In this paper, a RFWF feature selection method is proposed for building more effective feature subset in SER. A binary-tree structured multiclass and multiple-kernel SVM classifier is adopted to recognize emotions in a public emotional speech database. The experimental results indicate the effectiveness of the whole system.

The conclusions of this paper are as follows: (1) intrinsic properties of features about relevance, redundancy, and complementarity can be considered comprehensively by weights fusion. (2) Feature subset selected by RFWF achieves higher total accuracy than MRMR and DISR with lower dimension. (3) In multiclass classifier, different feature subsets adopted in different nodes can improve the recognizability of the whole system. (4) Multiple-kernel SVM classifier is robust and effective in recognizing the most confusable emotion.

The next work can focus on the research about more optimal feature fusion algorithms and automatic acquisition of optimal dimension of feature subset.

Conflicts of Interest

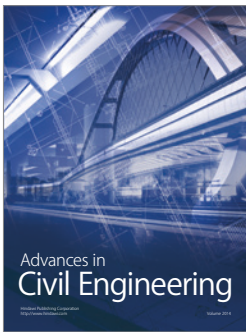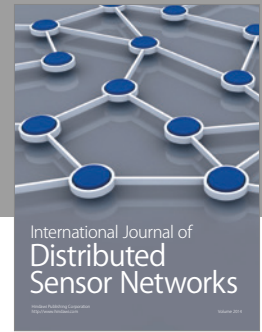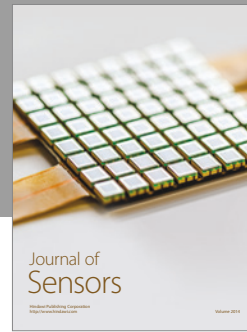The authors declare that they have no conflicts of interest.

References

[1] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, "Speech emotion recognition: features and classification models," *Digital Signal Processing*, vol. 22, no. 6, pp. 1154–1160, 2012.

[2] S. Chandaka, A. Chatterjee, and S. Munshi, "Support vector machines employing cross-correlation for emotional speech recognition," *Measurement: Journal of the International Measurement Confederation*, vol. 42, no. 4, pp. 611–618, 2009.

[3] C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, "Emotion recognition using a hierarchical binary decision tree approach," *Speech Communication*, vol. 53, no. 9-10, pp. 1162–1171, 2011.

[4] J. Yuan, L. Chen, T. Fan, and J. Jia, "Dimension reduction of speech emotion feature based on weighted linear discriminate analysis," in *Image Processing and Pattern Recognition*, vol. 8, pp. 299–308, International Journal of Signal Processing, 2015.

[5] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[6] H. Hua Yang and J. Moody, "Data visualization and feature selection: new algorithms for nongaussian data," *Advances in Neural Information Processing Systems*, vol. 12, pp. 687–693, 1999.

[7] D. Gharavian, M. Sheikhan, A. Nazerieh, and S. Garoucy, "Speech emotion recognition using FCBF feature selection method and GA-optimized fuzzy ARTMAP neural network," *Neural Computing and Applications*, vol. 21, no. 8, pp. 2115–2126, 2012.

[8] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[9] D. Lin and X. Tang, "Conditional infomax learning: an integrated framework for feature extraction and fusion," in *Proceedings of the 9th European Conference on Computer Vision*, pp. 68–82, Graz, Austria, 2006.

[10] P. E. Meyer, C. Schretter, and G. Bontempi, "Information-theoretic feature selection in microarray data using variable complementarity," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 261–274, 2008.

[11] X. Zhao, S. Zhang, and B. Lei, "Robust emotion recognition in noisy speech via sparse representation," *Neural Computing and Applications*, vol. 24, no. 7-8, pp. 1539–1553, 2014.

[12] A. Mencattini, E. Martinelli, G. Costantini et al., "Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure," *Knowledge-Based Systems*, vol. 63, pp. 68–81, 2014.

[13] D. Ververidis, C. Kotropoulos, and I. Pitas, "Automatic emotional speech classification," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. I-593–596, Quebec, CA, USA, 2004.

[14] J. Liu, C. Chen, J. Bu et al., "Speech emotion recognition based on a fusion of all-class and pairwise-class feature selection," in *Proceedings of the ICCS*, pp. 168–175, Beijing, China, 2007.

[15] X. Xu, Y. Li, X. Xu et al., "Survey on discriminative feature selection for speech emotion recognition," in *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing, ISCSLP 2014*, pp. 345–349, Singapore, Singapore, September 2014.

[16] L. Tian, X. Jiang, and Z. Hou, "Statistical study on the diversity of pitch parameters in multilingual speech," *Control and Decision*, vol. 20, no. 11, pp. 1311–1313, 2005.

[17] F. Burkhardt, A. Paeschke, M. Rolfes et al., "A database of German emotional speech," in *Proceedings of the 9th European Conference on Speech Communication and Technology*, pp. 1517–1520, Lisbon, Portugal, 2005.

[18] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing and Applications*, vol. 24, no. 1, pp. 175–186, 2014.

[19] C.-Y. Yeh, W.-P. Su, and S.-J. Lee, "An efficient multiple-kernel learning for pattern classification," *Expert Systems with Applications*, vol. 40, no. 9, pp. 3491–3499, 2013.

[20] G. R. G. Lanckriet, N. Cristianini, P. L. Bartlett et al., "Learning the kernel matrix with semidefinite programming," *Machine Learning Research*, vol. 5, no. 1, pp. 27–72, 2004.

[21] X. Jiang, K. Xia, X. Xia, and B. Zu, "Speech emotion recognition using semi-definite programming multiple-kernel SVM," *Journal of Beijing University of Posts and Telecommunications*, vol. 38, no. S1, pp. 67–71, 2015.

[22] B. Yang and M. Lugger, "Emotion recognition from speech signals using new harmony features," *Signal Processing*, vol. 90, no. 5, pp. 1415–1423, 2010.

[23] Y. Jin, P. Song, W. Zheng, and L. Zhao, "Novel feature fusion method for speech emotion recognition based on multiple kernel learning," *Journal of Southeast University*, vol. 29, no. 2, pp. 129–133, 2013.