Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 265819, 9 pages
http://dx.doi.org/10.1155/2013/265819

Research Article

Practical Speech Emotion Recognition Based on Online Learning: From Acted Data to Elicited Data

Chengwei Huang, Ruiyu Liang, Qingyun Wang, Ji Xi, Cheng Zha, and Li Zhao

School of Information Science and Engineering, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Chengwei Huang; huangcwx@126.com

Received 7 March 2013; Revised 26 May 2013; Accepted 4 June 2013

Academic Editor: Saeed Balochian

We study the cross-database speech emotion recognition based on online learning. How to apply a classifier trained on acted data to naturalistic data, such as elicited data, remains a major challenge in today's speech emotion recognition system. We introduce three types of different data sources: first, a basic speech emotion dataset which is collected from acted speech by professional actors and actresses; second, a speaker-independent data set which contains a large number of speakers; third, an elicited speech data set collected from a cognitive task. Acoustic features are extracted from emotional utterances and evaluated by using maximal information coefficient (MIC). A baseline valence and arousal classifier is designed based on Gaussian mixture models. Online training module is implemented by using AdaBoost. While the offline recognizer is trained on the acted data, the online testing data includes the speaker-independent data and the elicited data. Experimental results show that by introducing the online learning module our speech emotion recognition system can be better adapted to new data, which is an important character in real world applications.

1. Introduction

The state-of-the-art speech emotion recognition (SER) system is largely dependent on its training data. Emotional vocal behavior is personality dependent, situation dependent, and language dependent. Therefore, emotional models trained from a specific database may not fit to other databases. To solve this problem, we introduce an online learning framework to the SER system. Online speech data is used to retrain and to improve the classifier. Adopting the online learning framework, we may better adapt our SER system to different speakers and different data sources.

Many achievements have been reported on the acted speech emotion databases [1–3]. Tawari and Trivedi [4] considered the role of context and detected seven emotions on the Berlin Emotional Database [5]. Ververidis and Kotropoulos [6] studied gender-based speech emotion recognition system for five different emotional states. A number of machine learning algorithms have been studied in SER, using acted emotional data. Only recently the need of using naturalistic data has been pointed out. Several naturalistic speech emotion databases have been developed, such as AIBO emotional speech database [7] and VAM database [8]. Many researchers notice that real world data plays a key role in the SER system [9], and the model trained on the acted data does not fit very well on the naturalistic data.

Incremental learning may provide us a good solution to solve this problem under an online learning framework. The pretrained models on the acted data may be updated using very few online data. Since the naturalistic emotion data is very difficult to collect, acted speech data still plays an important role, especially in studying rare emotion types, such as fear-type emotion [1], confidence, and anxiety [10]. By using incremental learning we can make use of the available acted databases as a baseline recognizer and then retrain the classifier online for specific purposes.

Many successful algorithms have been proposed for incremental learning, such as Learning++ [11] and Bagging++ [12]. Incremental learning algorithms may be classified into two categories. In the first category, a single classifier is updated by reestimating its parameters. This type of learning algorithms is dependent on the specific classifier, such as the incremental learning algorithm for support vector machine proposed by Xiao et al. [13]. The techniques used in such parameter estimation may not be generalized. In the second category, the incremental learning algorithm is not dependent on a specific type of classifiers. Multiple classifiers are created and combined by a certain fusion rule, such as majority vote. Boosting is a typical type of algorithms that fall into the second category. By creating weak classifiers using selected data, we may add new training data to the learning procedure and gradually adapt the SER system in an online environment.

In this paper we explore the possibility of transferring pretrained SER system from acted data to more naturalistic data in an online learning framework. Section 2 describes our acted data and elicited data. Section 3 provides acoustic analysis of emotional features. In Section 4, we introduce our speech emotion recognizer and the online learning methodology. Finally, in Section 5, we provide the experimental results, which show that combining the acted data and the elicited data using online learning brings us the best result.

2. Three Types of Data Sources

In this paper we use three types of data sources to validate our SER system: (i) acted basic emotion database, (ii) speaker-independent emotion database, and (iii) elicited emotion database.

The first database contains the basic emotions, including happiness, anger, surprise, sadness, fear, and neutrality. The emotional speech is recorded by professional actors and actress, six males and six females. This acted database may be used as a standard training dataset for our baseline recognizer. However, in real world applications the naturalistic emotional speech is different from the acted speech.

The second database is designed for speaker-independent test, which includes fifty-one different speakers. Other than a large number of speakers, a special type of emotion is considered, namely, fidgetiness. Fidgetiness is an important type of emotion in cognitive related tasks. It may be induced by repeated work, environmental noise, and stress. The second database contains five emotions, as shown in Table 1. This database may be used for testing the ability of speaker adaptation. When using training data from the first database, it is challenging to test our SER system on the second database, due to many unknown speakers.

The third database contains elicited speech in a cognitive task, as shown in Table 2. The first row shows the emotion types collected in our experiments, such as fidgetiness, confidence, and tiredness. The second row is the speaker number related to each type of emotion. The third row is the male and female proportion in the emotion data. The last row is the number of utterances in each emotion class. The data is collected locally in our lab. We carried out a cognitive experiment and collected the emotional speech related to cognitive performance. Subject was required to work on a set of math calculations and to report the results orally. During the cognitive task the speech signals were recorded and annotated with emotional labels.

In the third database, "correct answer" or "false answer" labels are marked on each utterance in the oral report by the listeners who have not participated in the eliciting experiment. Therefore we may calculate the percentage of false answers in the negative emotion samples and the percentage of negative emotion in the "false answer" samples. Results show that the proportion of the mistake made in the math calculation is higher with the presence of negative emotions, as shown in Figures 1 and 2. The purpose of this database is to study the cognitive related emotions in speech. The analysis shows the dependency between the mistakes made in the math calculation and the negative emotions in the speech.

Table 1: The Speaker-independent emotion dataset.

Emotion type	Happiness	Anger	Fidgetiness	Sadness	Neutrality
Speaker number	51	51	51	51	51
Male/female	23/28	23/28	23/28	23/28	23/28
Utterance size	2200	2200	2200	2200	2200

Table 2: The Elicited Emotion Dataset.

Emotion type	Confidence	Tiredness	Fidgetiness	Happiness	Neutrality
Speaker number	6	6	6	6	6
Male/ female	3/3	3/3	3/3	3/3	3/3
Utterance size	1200	1200	1200	1200	1200

3. Feature Analysis

3.1. Acoustic Feature Extraction. Emotional information is hidden in the speech signals. Unlike the linguistic information, it is difficult to find the related acoustic features. Therefore feature analysis and selection are very important steps in building an SER system.

We selected typical utterances to study the feature variance caused by emotional change, as shown in Figures 3, 4, 5, 6, 7, 8, 9, 10, and 11. To better reflect the change caused by emotional information, we fix the context of these utterances.

The utterances shown in the figures are recorded from the same speaker. By comparing the utterances under different emotional state from the same speaker, we can exclude the influence brought by different speaking habits and personalities. It reveals the changes in the acoustic features caused only by the emotional information.

We induced three types of practical emotions from a cognitive task, namely, fidgetiness, confidence, and tiredness. We also studied the basic emotions, like happiness, anger, surprise, sadness, and fear. The intensity feature and the pitch contour are extracted and demonstrated in Figure 3 through Figure 11.

The first syllable is not normal speech under the fear emotional state. The pitch feature is missing, and it is whispered speech under the emotional state of fear. Under the tiredness emotion state, the pitch contour is low and flat, which is quite distinguishable from other emotion states.

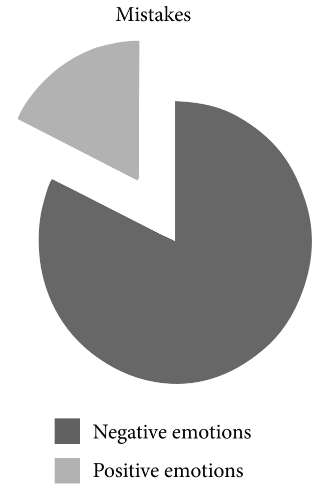


Figure 1: The percentage of negative emotions when mistake occurs in the cognitive task.

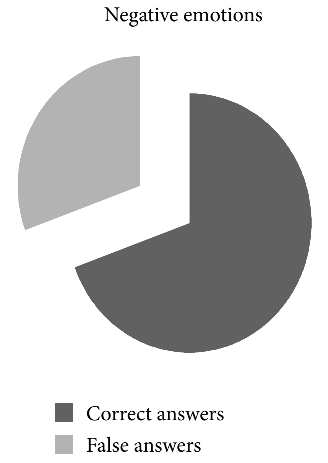


Figure 2: The percentage of correct answers and false answers when negative emotion occurs in the cognitive task.

In the neutral speech, the pitch contour is also flat, but at the end of the sentence the pitch frequency increases. Comparing speaking, the pitch frequency is not consistent at the end of the sentence. Under the sadness emotion state, the pitch contour is smooth and decreases at the end of the sentence. Furthermore, in the happiness sample, the variance of the pitch frequency is higher. The pith frequency also increases in the confidence and surprise samples.

We also notice that under the angry emotion state the variance of the intensity is lower and the intensity contour is smooth. However, in the sadness sample, the variance of the intensity is higher. Sadness and tiredness may have caused longer time duration and a lower speech rate, while fidgetiness and anger may have caused a higher speech rate.

Quantitative statistical analysis is shown in Figure 12. Pitch and formants features are compared under various emotional states.

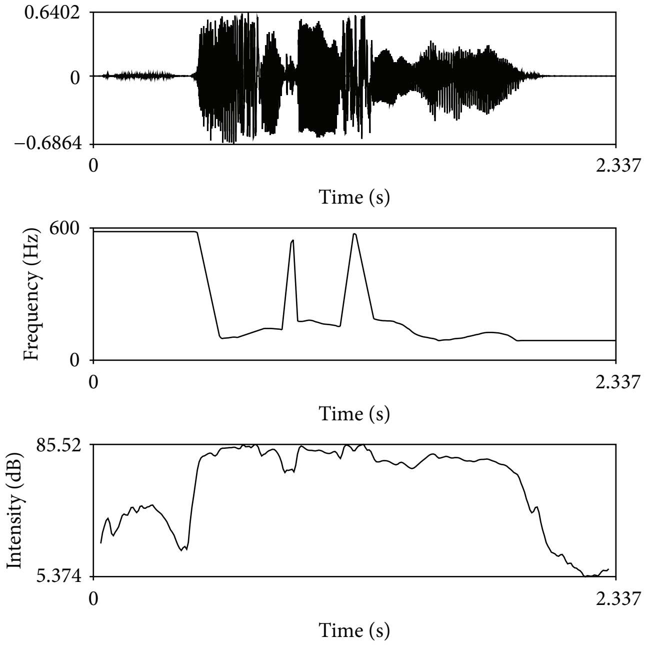


Figure 3: Intensity and pitch contour of happiness.

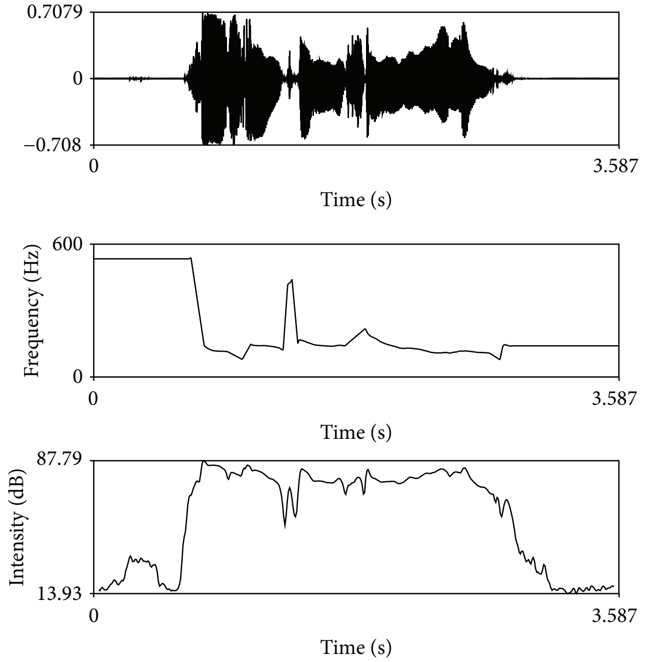


Figure 4: Intensity and pitch contour of sadness.

For modeling and recognition purposes, 481 dimensions of acoustic features are constructed. Statistic functions over the entire utterance, such as maximum, minimum, mean, range, are applied to the basic speech features, as listed below. "d" stands for difference and "d^2" stands for the second order of difference.

Feature 1–6: mean, maximum, minimum, median, range, and variance of Short-time Energy (SE).

Feature 7–18: mean, maximum, minimum, median, range, and variance of dSE and d^2SE.

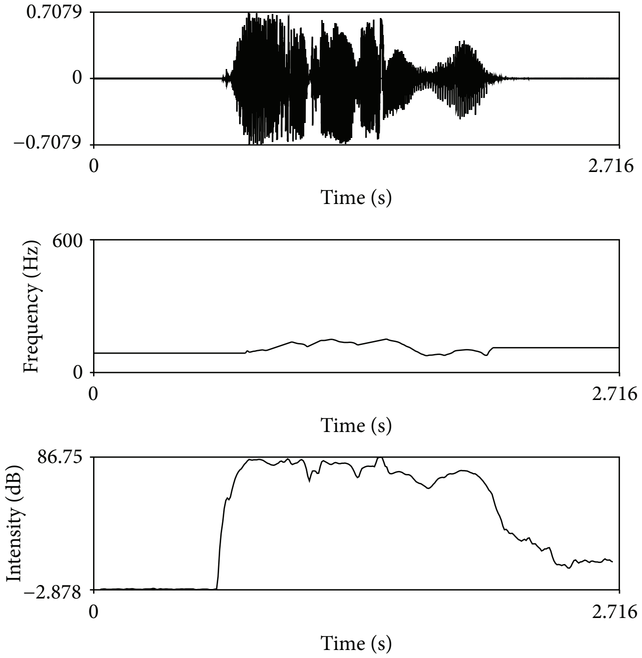


Figure 5: Intensity and pitch contour of fidgetiness.

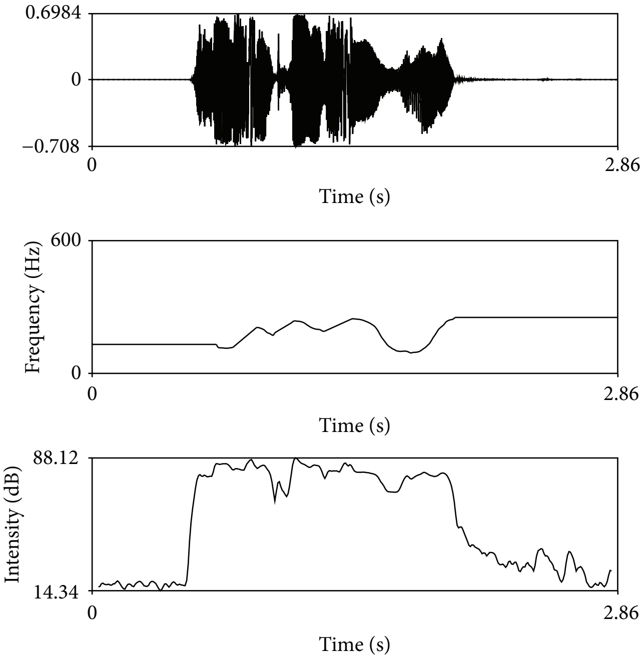


Figure 6: Intensity and pitch contour of surprise.

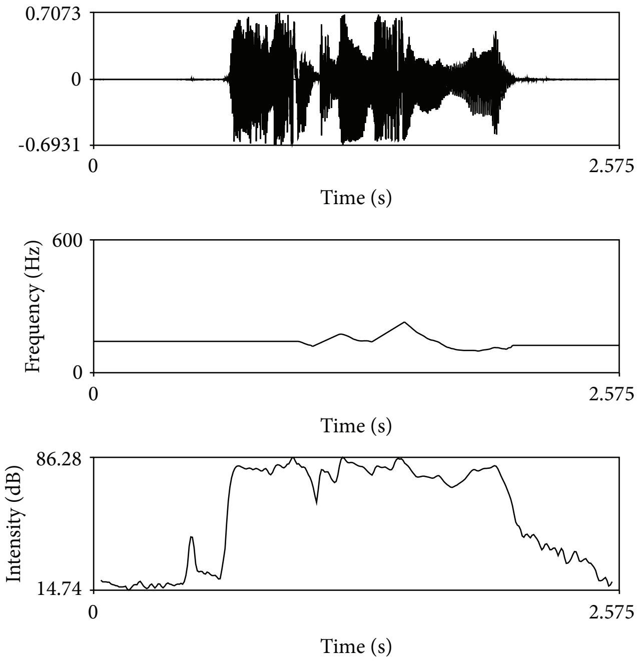


Figure 7: Intensity and pitch contour of fear.

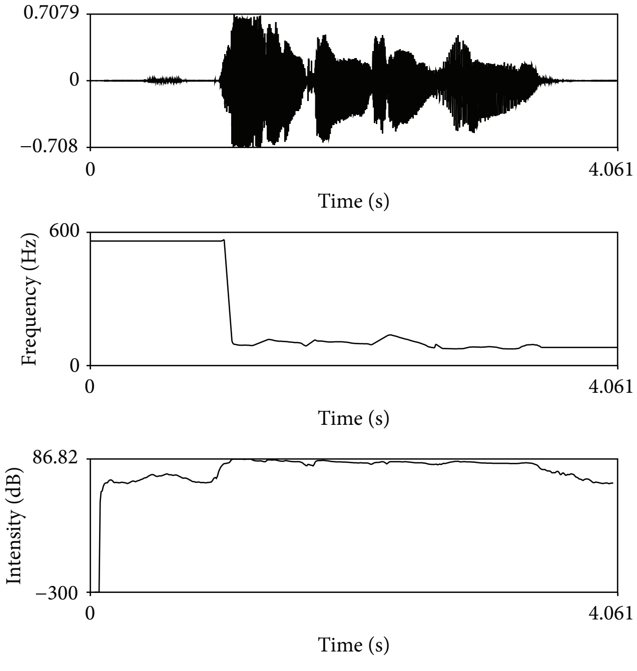


Figure 8: Intensity and pitch contour of tiredness.

Feature 19–24: mean, maximum, minimum, median, range, and variance of pitch frequency (F_0).

Feature 25–36: mean, maximum, minimum, median, range, and variance of dF_0 and d^2F_0.

Feature 37–42: mean, maximum, minimum, median, range, and variance of Zero-Crossing Rate (ZCR).

Feature 43–54: mean, maximum, minimum, median, range and variance of dZCR and d^2ZCR.

Feature 55: speech rate (SR).

Feature 56–57: Pitch Jitter1 (PJ1), Pitch Jitter2 (PJ2).

Feature 58–61: 0–250 Hz Energy Ratio (ER), 0–650 Hz ER, and 4 kHz above ER and Energy Shimmer (ESH).

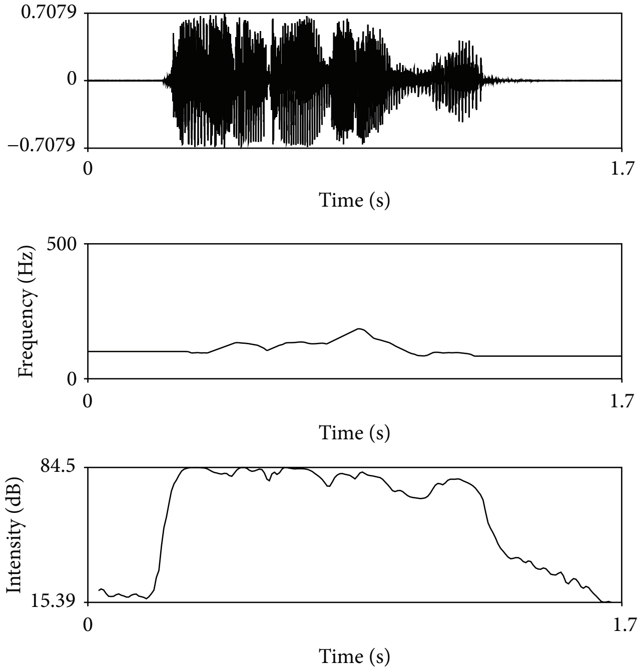


Figure 9: Intensity and pitch contour of anger.

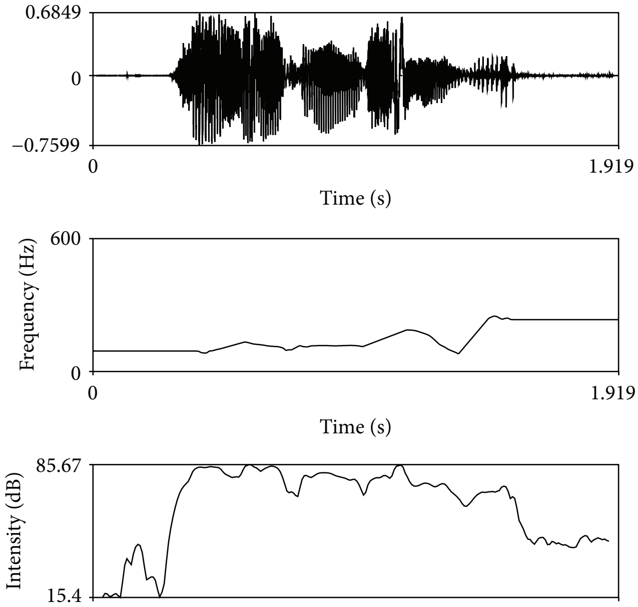


Figure 10: Intensity and pitch contour of neutrality.

Feature 62–65: Voiced Frames (VF), Unvoiced Frames (UF), UF/VF, and VF/(UF+VF).

Feature 66–69: Voiced Segments (VS), Unvoiced Segments (US), US/VS, and VS/(US+VS).

Feature 70-71: Maximum Voiced Duration (MVD), Maximum Unvoiced Duration (MUD).

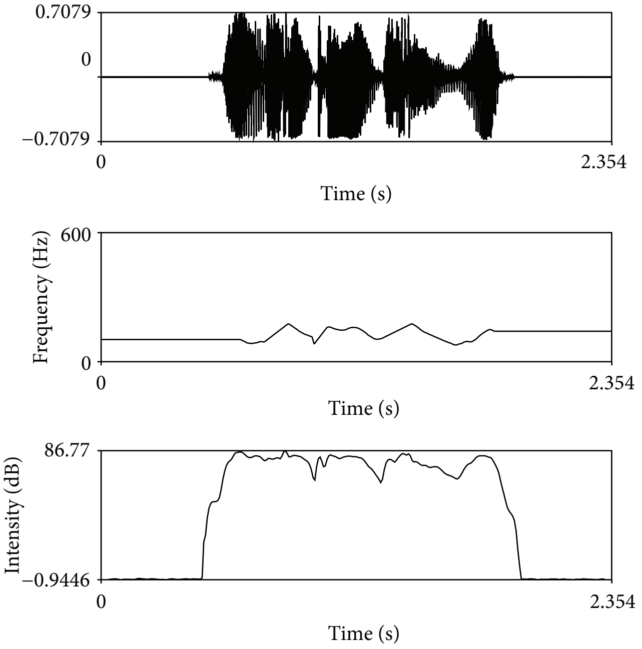


Figure 11: Intensity and pitch contour of confidence.

Feature 72–77: mean, maximum, minimum, median, range, and variance of Harmonic-to-Noise Ratio (HNR).

Feature 78–95: mean, maximum, minimum, median, range, and variance of HNR (0–400 Hz, 400–2000 Hz, and 2000–5000 Hz).

Feature 96–119: mean, maximum, minimum, median, range, and variance of 1st formant frequency (F1), 2nd formant frequency (F2), 3rd formant frequency (F3), and 4th formant frequency (F4).

Feature 120–143: mean, maximum, minimum, median, range, and variance of dF1, dF2, dF3, and dF4.

Feature 144–167: mean, maximum, minimum, median, range, and variance of d^2F1, d^2F2, d^2F3, and d^2F4.

Feature 168–171: Jitter1 of F1, F2, F3, and F4.

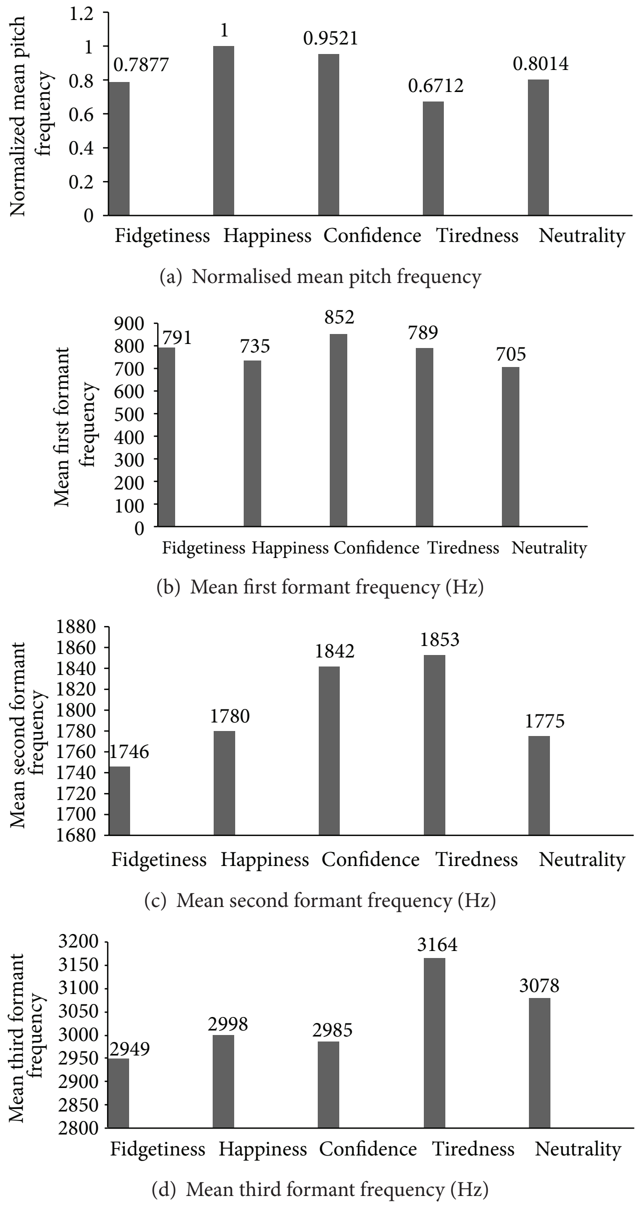
Feature 172–175: Jitter2 of F1, F2, F3, and F4.

Feature 176–199: mean, maximum, minimum, median, range, and variance of F1, F2, F3, and F4 Bandwidth.

Feature 200–223: mean, maximum, minimum, median, range, and variance of dF1 Bandwidth, dF2 Bandwidth, dF3 Bandwidth, and dF4 Bandwidth.

Feature 224–247 mean, maximum, minimum, median, range, and variance of d^2F1 Bandwidth, d^2F2 Bandwidth, d^2F3 Bandwidth and d^2F4 Bandwidth.

Feature 248–325: mean, maximum, minimum, median, range, and variance of MFCC (0–12th-order).



(a) Normalised mean pitch frequency

(b) Mean first formant frequency (Hz)

(c) Mean second formant frequency (Hz)

(d) Mean third formant frequency (Hz)

Figure 12: Feature distribution over various emotional states.

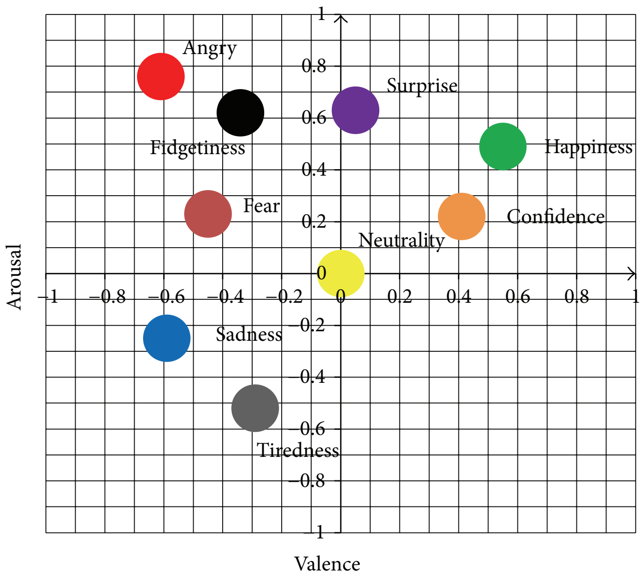


Figure 13: The arousal and the valence dimensions of emotions.

Feature 326–403: mean, maximum, minimum, median, range, and variance of dMFCC (0–12th-order).

Feature 404–481: mean, maximum, minimum, median, range, and variance of d2MFCC (0–12th-order).

3.2. Feature Selection Based on MIC. In this section we introduce the feature selection algorithm in our speech emotion classifier. Feature selection algorithms may be roughly classified into two groups, namely, “wrapper” and “filter.” Algorithms in the former group are dependent on the specific classifiers, such as sequential forward selection (SFS). The final selection result is dependent on a specific classifier. If we replace the specific classifier, the results will change. In the second group, feature selection is done by a certain evaluation criteria, such as Fisher Discriminant Ratio (FDR). The feature selection result achieved in this type of method is not dependent on specific classifiers and bears a better generality across different databases.

Maximal information coefficient (MIC) based feature selection algorithm falls into the second group. MIC is a new statistic tool that measures linear and nonlinear relationships between paired variables, invented by Reshef et al. [14].

MIC is based on the idea that if a relationship exists between two variables, then a grid can be drawn on the scatterplot of the two variables that partitions the data to encapsulate that relationship [14]. We may calculate the MIC of a certain acoustic feature and the emotional state by exploring all possible grids on the two variables. First, we compute for every pair of integers (x, y) that largest possible mutual information achieved by any x-by-y grid [14]. Second, for a fair comparison we normalize these MIC values between all acoustic features and the emotional state. Detailed study of MIC may be found in [14].

Since MIC can treat linear and nonlinear associations at the same time, we do not need to make any assumption on the distribution of our original features. Therefore it is especially suitable for evaluating a large number of emotional features. Based on a large number of basic features as described in Section 3.1, we apply MIC to measure the contribution of these features in correlation with emotion states. Finally a subset of features is selected for our emotion classifier.

4. Recognition Methodology

4.1. Baseline GMM Classifier. The Gaussian mixture model (GMM) based classifier is the state-of-the-art recognition method in speaker and language identification. In this paper we built the baseline classifier using Gaussian mixture model, and we may compare the baseline classifier with the online learning method.

GMM may be defined by the sum of several Gaussian distributions:

$$p\left(\mathbf{X}_t \mid \boldsymbol{\lambda}\right) = \sum_{i=1}^{M} a_i b_i\left(\mathbf{X}_t\right), \tag{1}$$

where $\mathbf{X}_t$ is a D-dimension random vector, $b_i(\mathbf{X}_t)$ is the ith member of Gaussian distribution, t is the index of utterance sample, a_i is the mixture weight, and M is the number of Gaussian mixture members. Each member is a D-dimension variable which follows the Gaussian distribution with the mean $\mathbf{U}_i$ and the covariance $\boldsymbol{\Sigma}_i$:

$$b_i\left(\mathbf{X}_t\right) = \frac{1}{(2\pi)^{D/2}\left|\boldsymbol{\Sigma}_i\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{X}_t - \mathbf{U}_i\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{X}_t - \mathbf{U}_i\right)\right\}. \tag{2}$$

Note that

$$\sum_{i=1}^{M} a_i = 1. \tag{3}$$

Emotion classification can be done by maximizing the posterior probability:

$$\text{EmotionLable} = \underset{k}{\operatorname{argmax}}\left(p\left(\mathbf{X}_t \mid \boldsymbol{\lambda}_k\right)\right). \tag{4}$$

Expectation Maximization (EM) is adopted for GMM parameter estimation [15]:

$$\begin{aligned}
a_m^i &= \frac{\sum_{t=1}^{T}\gamma_{tm}^i}{\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{tm}^i}, \\
\mathbf{U}_m^i &= \frac{\sum_{t=1}^{T}\gamma_{tm}^i \mathbf{X}_t}{\sum_{t=1}^{T}\gamma_{tm}^i}, \\
\boldsymbol{\Sigma}_m^i &= \frac{\sum_{t=1}^{T}\gamma_{tm}^i\left(\mathbf{X}_t - \mathbf{U}_m^i\right)\left(\mathbf{X}_t - \mathbf{U}_m^i\right)^T}{\sum_{t=1}^{T}\gamma_{tm}^i}, \\
\gamma_{tm}^i &= \frac{a_m^{i-1} N\left(\mathbf{X}_t \mid \mathbf{U}_m^{i-1}, \boldsymbol{\Sigma}_m^{i-1}\right)}{\sum_{m=1}^{M} a_m^{i-1} N\left(\mathbf{X}_t \mid \mathbf{U}_m^{i-1}, \boldsymbol{\Sigma}_m^{i-1}\right)}.
\end{aligned} \tag{5}$$

Due to the different types of emotions among the datasets, we unify the emotional datasets by categorizing them into positive and negative regions in the valence and arousal dimensions, as shown in Figure 13. We may verify the ability of the emotion classifier by classifying the emotional utterances into different regions in the valence and arousal space.

4.2. Online Learning Using AdaBoost. While the offline GMM classifier is trained using EM algorithm, the online training algorithm using AdabBoost will be introduced in this section. AdaBoost is a powerful algorithm in assemble learning [16]. The belief in this AdaBoost is that weak classifiers may be combined into a powerful classifier. Multiple classifiers trained on randomly selected datasets perform quiet differently from each other on the same testing dataset; therefore, we may reduce the misclassification rate by a proper decision fusion rule.

AdaBoost algorithm consists of several iterations. In each iteration, a new training set is selected for a new weak classifier. A weight is assigned to the new weak classifier. Based on the testing results of the new weak classifier, the weights of all the data samples are modified for the next iteration. At the final step the assembled classifier is achieved by combination of the multiple weak classified through a weighted voting rule.

Let us suppose the current training set is [17]

$$T = \left\{s_1, s_2, \ldots, s_N\right\}, \tag{6}$$

where the weights of the samples are

$$\begin{gathered}
W = \left\{w_1, w_2, \ldots, w_N\right\}, \\
\sum_{i=0}^{N} w_i = 1.
\end{gathered} \tag{7}$$

The error rate of the new weak classifier is

$$e = \sum_{i:c(s_i) \neq y_i} w_i, \tag{8}$$

where $c(s_i)$ is the classification result and y_i is the class label. The fusion weight assigned to each classifier is defined by the error rate:

$$\alpha = \ln\left(\frac{(1-e)}{e}\right). \tag{9}$$

At the beginning of the algorithm, each sample is assigned by equal weight. During the iteration, the sample weights are updated:

$$w_{i+1} = \begin{cases} w_i \times \beta, & c\left(s_i\right) \neq y_i, \\ w_i, & c\left(s_i\right) = y_i. \end{cases} \tag{10}$$

At the arrival of the new data, assuming that we know the label information for each sample, pretrained classifiers from the offline data are used as initial weak classifiers. AdaBoost algorithm is applied to the new online data, and fusion weights are reassigned to the offline trained classifiers.

At the first m initial iterations, m pretrained classifiers are used as the weak classifiers and added to the final ensemble classifier, instead of training new weak classifiers from the randomly selected dataset. After the m initial iterations, new weak classifiers are trained from the new online data and added to the final ensemble classifier in the AdaBoost algorithm.

The major difference between the online training and the offline training is the data used for learning. Offline training uses large acted data, while online training uses small and natural data. Offline training is independent of the online training and ready to use, while the online training is dependent on the offline training and only retrains the existing model to fit specific purposes, such as to tune on a large number of speakers. The purpose of online training is to quickly adapt the existing offline model to a small amount of new data.

5. Experimental Results

In our experiment, the offline training is carried out on the acted basic emotion dataset. The speaker-independent dataset and the elicited practical emotion dataset are used for the online training and the online testing. Although the datasets used in online testing are preprocessed utterances rather than real time online data, our experiments still provide a simulated online situation. We divide dataset 2 and dataset 3 into smaller sets, dataset 2a and dataset 2b, which are used as the simulated online initialization.

Speech utterances from different sources are organized into several datasets, as shown in Table 2.

The online learning algorithm is verified both on the speaker-independent data and the elicited data. The results are shown in Table 4. A large number of speakers bring difficulties in modeling emotional behavior, since emotion expression is highly dependent on individual habit and personality. By extending the offline trained classifier to the online data that contains a large number of speakers, we improved the generality of our SER system. The elicited data is collected in a cognitive experiment that is more close to the real world situation. During the cognitive task emotional speech is induced. We observed that the different nature between the acted data and the induced speech during a cognitive task caused a significant decrease of the recognition rate. By using the online training technique we may transfer the offline trained SER system to the elicited data. Extending our SER system to different data sources may bring emotion recognition closer to real world applications.

The major challenge in our online learning algorithm is how to combine the existing offline classifier and efficiently adapt the model parameters to a small number of new online data. We adopted the incremental learning idea and solved this problem by modifying the initial stage in the AdaBoost framework. One of the contributions of our online learning algorithm is that we may reuse the existing offline training data and make the online learning stage more efficiently. We make use of a large amount of available offline training data and only require a small amount of data for online training, as shown in Table 3. The weight of each weak classifier is an important parameter. The proposed method may be further improved by using fuzzy membership function to evaluate the confidence in GMM classifiers and reestimate the weight of each weak classifier.

6. Discussions

Acted data is often considered not suitable for real world applications. However, traditional researches have been focused on the acted emotion speech, and many acted databases are available. How to transfer an SER system that trained on the acted data to the new naturalistic data in real world is an unsolved challenge.

Many feature selection algorithms may be applied to SER system. MIC is a newly proposed and powerful algorithm for exploring nonlinear relationship between variables.

AdaBoost is a popular algorithm to ensemble multiple weak classifiers to establish a strong classifier. By applying AdaBoost in the online occasion, we train multiple weak classifiers based on the newly arrived online data. The offline pretrained classifiers are used for initialization. We may explore other incremental learning algorithms in the future work.

Table 3: Selected datasets for online and offline experiments.

Datasets index	Data source	Number of utterances	Purpose of use
Dataset 1	Acted speech	12000	Offline training
Dataset 2a	Speaker independent	1000	Online training
Dataset 2b	Speaker independent	10000	Testing
Dataset 3a	Elicited speech	1000	Online training
Dataset 3b	Elicited speech	5000	Testing

Table 4: Online and offline experimental results.

Experiment index	Offline training set	Online training set	Testing set	Classification result %
Experiment 1	Dataset 1	N/A	Dataset 2b	63.3%
Experiment 2	Dataset 1	Dataset 2a	Dataset 2b	75.6%
Experiment 5	Dataset 2a	N/A	Dataset 2b	70.0%
Experiment 3	Dataset 1	N/A	Dataset 3b	61.2%
Experiment 4	Dataset 1	Dataset 3a	Dataset 3b	73.1%
Experiment 6	Dataset 3a	N/A	Dataset 3b	68.5%

Acknowledgments

This work was partially supported by China Postdoctoral Science Foundation (no. 2012M520973), National Nature Science Foundation (no. 61231002, no. 61273266, no. 51075068), and Doctoral Fund of Ministry of Education of China (no. 20110092130004). The authors would like to thank the anonymous reviewers for their valuable comments and helpful suggestions.

References

[1] C. Clavel, I. Vasilescu, L. Devillers, G. Richard, and T. Ehrette, “Fear-type emotion recognition for future audio-based surveillance systems,” *Speech Communication*, vol. 50, no. 6, pp. 487–503, 2008.

[2] C. Huang, Y. Jin, Y. Zhao, Y. Yu, and L. Zhao, “Speech emotion recognition based on re-composition of two-class classifiers,” in *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII ’09)*, Amsterdam, The Netherlands, September 2009.

[3] K. R. Scherer, “Vocal communication of emotion: a review of research paradigms,” *Speech Communication*, vol. 40, no. 1-2, pp. 227–256, 2003.

[4] A. Tawari and M. M. Trivedi, “Speech emotion analysis: exploring the role of context,” *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 502–509, 2010.

[5] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, “A database of German emotional speech,” in *Proceedings*

of the 9th European Conference on Speech Communication and Technology, pp. 1517–1520, Lissabon, Portugal, September 2005.

[6] D. Ververidis and C. Kotropoulos, "Automatic speech classification to five emotional states based on gender information," in *Proceedings of the 12th European Signal Processing Conference*, pp. 341–344, Vienna, Austria, 2004.

[7] S. Steidl, *Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech*, Department of Computer Science, Friedrich-Alexander-Universitaet Erlangen-Nuermberg, Berlin, Germany, 2008.

[8] M. Grimm, K. Kroschel, and S. Narayanan, "The Vera am Mittag German audio-visual emotional speech database," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '08)*, pp. 865–868, Hannover, Germany, June 2008.

[9] K. P. Truong, *How Does Real Affect Affect Affect Recognition in Speech?* Center for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, 2009.

[10] C. Huang, Y. Jin, Y. Zhao, Y. Yu, and L. Zhao, "Recognition of practical emotion from elicited speech," in *Proceedings of the 1st International Conference on Information Science and Engineering (ICISE '09)*, pp. 639–642, Nanjing, China, December 2009.

[11] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, "Learn++: an incremental learning algorithm for supervised neural networks," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 31, no. 4, pp. 497–508, 2001.

[12] Q. L. Zhao, Y. H. Jiang, and M. Xu, "Incremental learning by heterogeneous Bagging ensemble," *Lecture Notes in Computer Science*, vol. 6441, no. 2, pp. 1–12, 2010.

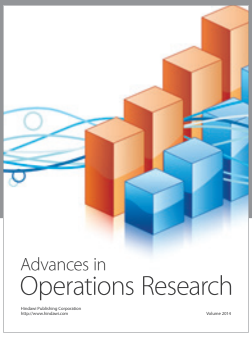
[13] R. Xiao, J. Wang, and F. Zhang, "An approach to incremental SVM learning algorithm," in *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, pp. 268–273, 2000.

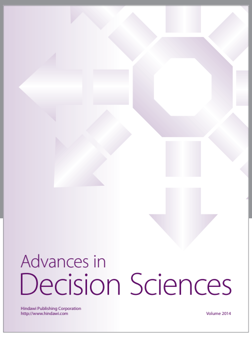
[14] D. N. Reshef, Y. A. Reshef, H. K. Finucane et al., "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.

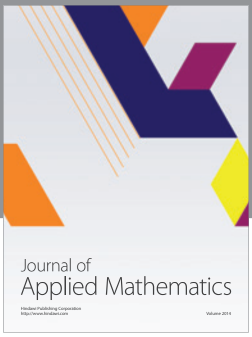
[15] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.

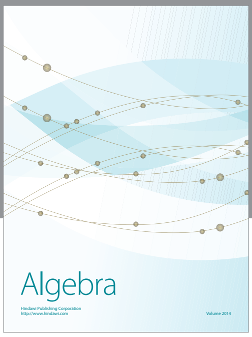
[16] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, part 2, pp. 119–139, 1997.

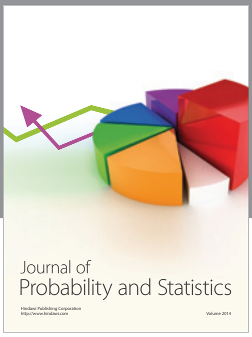
[17] Q. Zhao, *The research on ensemble pruning and its application in on-line machine learning [Ph.D. thesis]*, National University of Defense Technology, Changsha, China, 2010.


Advances in
Operations Research


Advances in
Decision Sciences

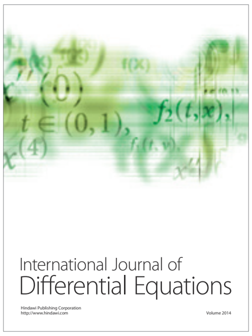

Journal of
Applied Mathematics

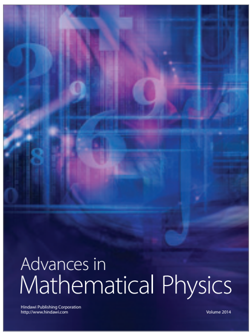

Algebra

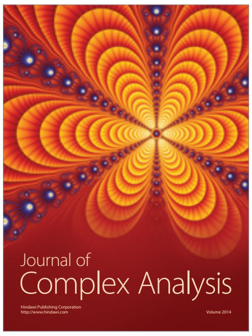

Journal of
Probability and Statistics

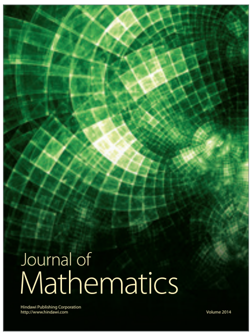
The Scientific
World Journal

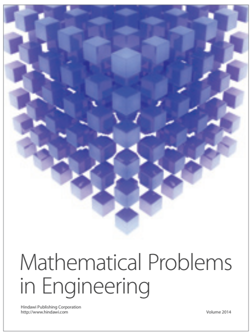
International Journal of
Combinatorics

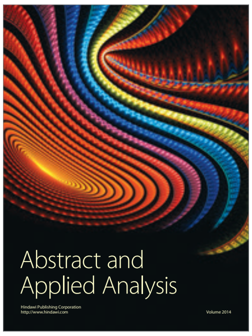

International Journal of
Differential Equations

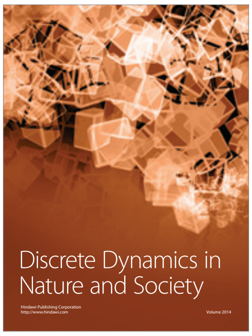

Advances in
Mathematical Physics

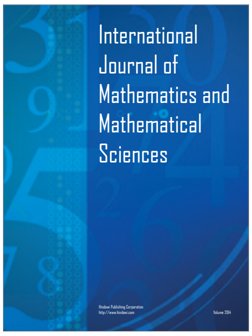

Journal of
Complex Analysis


Journal of
Mathematics


Mathematical Problems
in Engineering

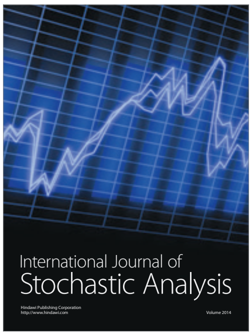

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


International Journal of
Stochastic Analysis

Journal of
Optimization