Resbee Publishers

Multimedia Research (MR)

Received 20 September, Revised 28 November, Accepted 30 December

Impact of Supervised Classifier on Speech Emotion Recognition

Anitha J.S

Department of Computer Scirnce and Engineering,
SCAD College of Engineering and Technology
Tirunelveli, Tamil Nadu, India
ajooani@gmail.com

Abinaya J.S

Department of Communication and networking,
CSI Institute of Technology
Kanyakumari, Tamil Nadu, India
meabianu@gmail.com

Abstract: A face recognition system is a computer application proficient of verifying or identifying a person from a video frame or a digital image from a video source. The human face acts a significant role in the social communication, passing on people's uniqueness. By means of the human face as a key to protection, biometric face recognition technology has attained noteworthy consideration in the precedent numerous years owing to its prospective for an extensive assortment of applications in both non-law enforcement and law enforcement activities. In this paper, the Speech Emotion Recognition (SER) is analyzed by adopting cepstral features for feature extraction and k-NN classifier for classification. Moreover, the implemented process is compared with k-means and C-means algorithms and the results are obtained.

Keywords: Speech Emotiuon Recognition; k-NN; Cepstral features; Accuracy; Database

1. Introduction

In current years, technology for automatic emotion recognition from speech has developed adequately to be concerned in numerous real-life scenarios, namely remote education, call centres, driving safety, computer games, and auxiliary disease diagnosis. Nevertheless, conventional SER technology has not accomplished very excellent presentation, perhaps owing of the deficient of effectual emotion-associated features.SER is a significant exploration field in the understanding of Artificial Intelligence (AI). Noise present in the signal processing and systems environment manipulates the recognition precision and confines the realistic applications of SER, like, adjuvant therapy systems for autism and intelligent customer service systems, in which precise emotion recognition is essential to formulate a appropriate response.

Feature selection is an essential measure in the improvement of a system for recognizing emotions in speech. In recent times, the communication among features produced from the similar audio source was hardly ever measured that may perhaps generate redundant features and raise the performance costs. In SER, an essential research problem is how to choose a best possible feature set from speech signals [6]. Numerous traditional state-of-the-art speech emotion recognition methods usually assume that the features of the training and test samples are strained from the similar distribution. This postulation does not embrace in numerous real world applications. This is primarily owing to the speech signals from dissimilar provinces that are very contradictory regarding type of emotion, speakers, degree of spontaneity and recording circumstances. A classifier just trained on a particular corpus and subsequently pertained directly to a different corpus, which cannot be expected to include a tremendous performance.

The majority of the preceding works on SER have been dedicated on the investigation of spectral information and speech prosodic characteristics. And several new feature constraints are adopted for SER, like the Fourier parameters. Even though there are numerous acoustic parameters which have been demonstrated to enclose emotional information, only a small accomplishment has been attained in understanding such a set of characteristics that consistently executes over dissimilar circumstances [9]. Therefore, the majority of researchers desire to exploit mixing feature set which is comprised of several kinds of features including further emotional information [10]. On the other hand, exploitation of mixing feature set may perhaps increase the high redundancy and dimension of speech features, thus it formulates the learning process complex for the majority of machine learning schemes and raises the

likelihood of overfitting. Consequently, feature selection is indispensible to decrease the redundancy in dimensions of features. SER can be considered as a static or dynamic classification crisis that makes SER an excellent test bed for exploring and adopting a variety of deep learning architectures. However, for the significance of SER technology, the majority of them are dependent on the application surroundings of small- sample and small scale.

This paper contributes the emotion recognition from speech using k-NN algorithm for cepstral features. Moreover, the capability of SER is validated by means of various performance measures such as accuracy, sensitivity, and specificity, precision, FPR, FNR, NPV, FDR, F1-score and MCC and the improvements of the proposed scheme is verified. This paper is organized as follows. Section II describes the related works and reviews done under this topic. Section III explains the SER modelling and section IV demonstrates the results and discussions. Finally, section V concludes the paper.

2.Literature Review

2.1 Related Works

In 2018, ShaolingJing *et al.* [1], has suggested a novel kind of feature associated to prominence, in concert with conventional acoustic characteristics were deployed to categorize seven distinctive emotional states. Moreover, the author group generated a Chinese Dual-mode Emotional Speech Database (CDESD) that includes paralinguistic annotation and supplementary prominence information. Subsequently, a consistency assessment scheme was presented to authenticate the reliability of the annotation details of such database. The consequences demonstrate that the annotation constancy on prominence accomplishes more than 60% on average. Consequently, this research investigates the correlation of the prominence characteristics with emotional states by means of a curve fitting technique. Prominence was established to be intimately associated to emotion states, to maintain emotional information to the maximum feasible amount and to act a significant role in emotional expression. At last, the suggested prominence features were authenticated on CDESD via speaker-independent and speaker-dependent researches with four generally adopted classifiers.

In 2017, Qirong *et al.* [2], has implemented an orthogonal variable to persuade the input to be extricated into two blocks: emotion-unrelated and emotion-related features. The implemented technique can be trained with domain invariant and emotion-discriminative features by using a back propagation network that deploys the acoustic characteristics of INTERSPEECH 2009 Emotion Challenge as the input more willingly than raw speech signals. Experimentations performed on the INTERSPEECH 2009 Emotion Challenge two-class task demonstrates that the computation of the suggested process was advanced when compared with other state-of-the-arts techniques.

In 2017, Haytham *et al.* [3], has introduced a frame-dependent formulation to SER, which was based on end-to-end deep learning and minimal speech processing to design intra-utterance dynamics. Moreover, the implemented SER system was proposed to empirically discover neural network and feed-forward recurrent variants and their architectures. Experimentations carried out elucidate the advantages and restrictions of these frame works in paralinguistic emotion recognition and speech recognition in specific. As a consequence of the investigation, conventional outcomes on the IEMOCAP database were explored for present quantitative and qualitative assessments and speaker-independent SER of the performance models.

In 2017, Jiang *et al.* [4] has established the reconstruction of speech samples that eliminates the noise that was added. Acoustic features obtained from the reconstructed samples were chosen to construct an optimal feature subset with enhanced recognisability of emotions. A multiple-kernel (MK) support vector machine (SVM) classifier which was resolved by semi-definite programming (SDP) was deployed in SER course of action. The implemented technique in this paper was established on Berlin Database of Emotional Speech. Recognition precision of the noisy, original, and reconstructed samples categorized by both single-kernel (SK) and MK classifiers were categorized and analyzed. The investigational results demonstrate that the suggested process was robust and effectual when noise subsists.

In 2018, Zhen *et al.* [5] has implemented an emotion recognition technique depending on Extreme Learning Machine (ELM) decision tree that was suggested based on the confusion degree between diverse fundamental emotions. A structural design of speech emotion recognition was implemented and the classification researches depending on presented classification technique by means of Chinese speech database from institute of automation of Chinese academy of sciences (CASIA) were carried out. In addition, the investigational results demonstrate that this method has attained 89.6% recognition rate on common. Accordingly, it would be efficient and fast to differentiate emotional states of diverse

speakers from speech, and it would formulate it feasible to understand the interaction among computer/robot and speaker-independent in the upcoming generations.

2.2 Review

Table 1 shows the methods, features and challenges of conventional techniques based on wind generation of IPMSG using MPPT algorithm. At first, Consistency assessment algorithm was proposed in [1] which offers Average recognition rate with better reliability. However, it was difficult to determine the prominence trends throughout visual assessment. Moreover, Domain-invariant Feature Learning Method (EDFLM) was suggested in [2] that provide high level of statistical significance and also the reproducibility was ensured, but there was no contemplation on reducing discrepancy between the source and target domains. In addition, Convolution Neural network (CNN) was established in [3], which does not depend on future context and it was able to contract with utterances of arbitrary length with no deprivation. However, there was no integration of SER system with automatic speech recognition. Similarly, Multiple-kernel (MK) Support Vector Machine (SVM) was suggested in [4] that provide better effective process with increased robustness; anyhow, it necessitates faster implementation of sample reconstruction and SDP solving. Finally, decision tree was established in [5] that offers minimized time complexity with improved validity of the feature selection. However, there was no feature selection based on evolutionary computation. These above mentioned challenges are considered for motivating the improvement of SER.

Table 1: *Review on conventional SER systems*

Author [citation]	Adopted methodology	Features	Challenges
ShaolingJing *et al.* [1]	Consistency assessment algorithm.	Average recognition rate Better reliability	Difficult to discover the prominence trends through visual inspection
Qirong *et al.* [2]	Domain-invariant Feature Learning Method (EDFLM)	High level of statistical significance Reproducibility is ensured	No contemplation on reducing discrepancy between the source and target domains
Haytham *et al.* [3]	Convolution Neural network (CNN)	Does not depend on future context Able to contract with utterances of arbitrary length with no deprivation	No integration of SER system with automatic speech recognition
Jiang *et al.* [4]	Multiple-kernel (MK) support vector machine (SVM)	Offers better effective process Provides increased robustness	Requires faster Implementation of sample reconstruction and SDP solving.
Zhen *et al.* [5]	Decision tree	Reduced time complexity. Better validity of the feature selection	No feature selection based on evolutionary computation

3. Speech Emotion Recognition Modelling

3.1 Proposed Speech Emotion Design

The overall architecture of the implemented SER model is described by Fig. 1. Initially, the samples of speech signals are given as input for extracting the features. In this paper, Cepstral feature extraction technique is adopted to extract the features in a better way. The extracted features are then subjected to classification by means of k-NN classifier, from which the speech signals are recognized and obtained.

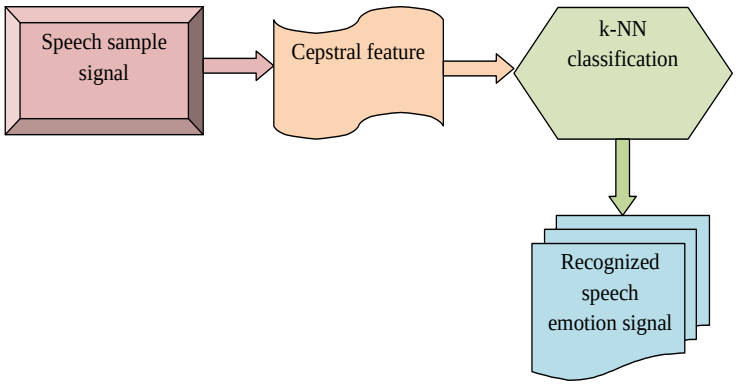

Fig. 1. *Overall framework of the proposed model*

3.2 Cepstral Features

Consider that the speech signal be $p(n)$, that is attained from the convolution of two signals $f(n)$ and $i(n)$ as the computation of two signals and it is given by Eq. (1), in which $\hat{p}(n)$ is the complicated cepstrum [19]. $\hat{p}(n)$ is indicated as given in Eq. (8), in which, $i(n)$ and $f(n)$ are the silent portion and speech portion of the speech recording correspondingly. The cepstral analysis depends on the subsequent observation that is specified in Eq.(2). The log of the signal $S(z)$ is described by Eq. (3)

$$p(n) = f(n) + i(n) \rightarrow \hat{p}(n) = \hat{f}(n) + \hat{i}(n) \tag{1}$$

$$p(n) = p_1(n) * p_2(n) \Leftrightarrow P(z) = P_1(z)P_2(z) \tag{2}$$

$$\log\{P(z)\} = \log\{P_1(z)\} + \log\{P_2(z)\} = \hat{P}(z) \tag{3}$$

If the Z transform is suitable and the complex log is typical, then both the convolved signals $\hat{p}_1(n)$ and $\hat{p}_2(n)$ are additive as revealed by Eq.(4). The signal $p(n)$ is constrained to have poles and zeros into the unit circle, that is denoted as in Eq. (5), in which $\log\{P(v)\}$ is the complex logarithm of $P(v)$.

$$\hat{p}(n) = \hat{p}_1(n) + \hat{p}_2(n) \tag{4}$$

$$\log\{P(v)\} = \log\{|P(v)|\} + l \sphericalangle P(v) \tag{5}$$

If $P(v) = P_1(v)P_2(v)$ then $\log\{|P(v)\}$ is indicated as exposed in Eq. (6). The real cepstrum $R_x(n)$ is described as specified in Eq. (7), in which the magnitude of $R_x(n)$ is real in addition to non-negative. The complex cepstrum $\hat{p}(n)$ is described in Eq.(8), in which the phase is indicated as $\arg(\ )$, $\log|P(e^{lv})|$ and $\log\{|P(v)\}$ indicates the log spectrum of the signal. This is multifaceted since it exploits the complex logarithm. In addition, the composite cepstrum of the real sequence is real.

$$\log\{|P(v)|\} = \log\{|P_1(v)P_2(v)\} = \log\{|P_1(v)|\} + \log\{|P_2(v)|\} \tag{6}$$

$$R_x(n) = \frac{1}{2\Pi}\int_{-\Pi}^{\Pi} \log|P(e^{lv})|\, e^{lvn}\, dv \tag{7}$$

$$\hat{p}(n) = \frac{1}{2\Pi}\left[\left[\log|P(e^{lv})|\right] + l\arg(P(e^{lv}))\right]e^{lvn}\, dv \tag{8}$$

Actually, the real cepstrum is the even part of $\hat{p}(n)$, and it is illustrated in Eq.(9). $R_x(n)$, is usually exploited in the speech processing, that is obatined by deploying an Inverse Fourier Transform (IFT) of the log spectrum of the signal.

$$R_x(n) = \frac{\hat{p}(n) + \hat{p}(-n)}{2} \tag{9}$$

In digital signals, the Fourier transform is substituted by Discrete Fourier transform (DFT), and it is revealed in Eq.(10), in which $\hat{P}_t(h)$ is indicated as the sampled version of $\hat{X}(e^{jw})$ and thus $\hat{p}_t(n)$ is explained as revealed in Eq.(11), in which N indicates the period. In the same way, aliasing by recurrence of cepstrum with N, the cepstral feature $R_t(n)$ of $p(n)$ is potrayed as specified in Eq. (12).

$$\begin{aligned} P_t(h) &= \sum_{n=0}^{N-1} p(n) e^{-l\frac{2\Pi}{N}hn} \qquad 0 \le h \le N-1 \\ \hat{P}_t(h) &= \log\{P_t(h)\} \qquad 0 \le h \le N-1 \\ \hat{p}_t(n) &= \frac{1}{N}\sum_{h=0}^{N-1} \hat{P}_t(h) e^{l\frac{2\Pi}{N}hn} \quad 0 \le n \le N-1 \end{aligned} \tag{10}$$

$$\hat{p}_t(n) = \sum_{p=-\infty}^{\infty} \hat{p}(n + qN) \tag{11}$$

$$R_t(n) = \frac{1}{N}\sum_{h=0}^{N-1} \log|R_t(h)|\, e^{l\frac{2\Pi}{N}hn} \quad 0 \le n \le N-1 \tag{12}$$

3.3 K-NN Classification

k-NN classification is a renowned decision rule that is extensively exploited in classification of patterns [20].

Consider l be the count of classes, and $\mu\Delta\left\{d^{(i)}, i = 1,2,...l\right\}$ be the group of class labels. Assume that γ be a group of labelled patterns, which are denoted to as a template. A labelled pattern $y \in \Re^n$ in the template is indicated as a prototype, in which n signifies the dimension of patterns. The variable $v(y)$ represents the weight of a prototype y, that is the count of prototypes y in the template. The class label of a prototype y is indicated by $d(y)$.

Assume that the matching computation among pattern x and y be $\lambda(x,y)$, in which λ is considered to be a symmetric and non-negative function. The higher value of $\lambda(x,y)$ demonstrates the improved degree of relationship among x and y. The reciprocal of the Hamming distance is a generally utilized assess for binary pattern matching as given by Eq. (13)

$$\lambda(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|\right)^{-1} \tag{13}$$

Let $y^{(1)}, y^{(2)} ... y^{(d)}$, be the d prototypes that are adjacent to x, in the sense of λ between each and every prototypes in the template γ, and in addition, it gratifies $\sum_{j=1}^{d} v\left(y^{(j)}\right) = k$. The unweighted voting power of all the classes are evaluated as given by Eq. (14), in which $\delta(.,.)$ gratifies Eq. (15).

$$\beta_i \Delta \sum_{j=1}^{d} \delta\left(y^{(j)}, d^{(i)}\right) v\left(y^{(j)}\right) \quad i = 1,2,...l \tag{14}$$

$$\delta\left(y^{(j)}, d^{(i)}\right) \Delta \begin{cases} 1 & \text{if } d^{(i)} = d\left(y^{(i)}\right) \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

$$\tilde{\beta}_i \Delta \sum_{j=1}^{d} \lambda\left(x, y^{(j)} d^{(i)}\right) \delta\left(y^{(j)}, d^{(j)}\right) v\left(y^{(j)}\right) \quad i = 1,2,...l \tag{16}$$

Moreover, the weighted voting power of all the classes can be evaluated as given by Eq. (16).

4. Results And Discussions

4.1 Simulation Procedure

The proposed k-NN model was experimented in MATLAB and the results were obtained. Three databases were adopted for experimentation, namely, benchmark database, Marathi database and Hindi database. The proposed k-NN model was compared with k-means [20] and C-means [21] algorithms with respect to various measures such as, accuracy, sensitivity, specificity, precision, FPR, FNR, NPV, FDR, F1-score and MCC and the improvements of the suggested scheme were verified.

4.2 Performance Analysis

The performance analysis of the proposed k-NN classifier was compared with k-mean and C-mean for cepstral features. The Marathi database can be accessed from Table II, in which the suggested scheme regarding accuracy is 9.5% better than k-means and 2.56% better than C-means methods. Similarly, the suggested scheme regarding sensitivity is 16.7% superior to k-means and 12.37% superior to C-means methods. In addition, the specificity of the implemented scheme is 14.65% better than k-means and 15.22% better than C-means techniques. Moreover, the precision of proposed scheme is 11.87% superior to k-means and 4.35% superior to C-means methods. Also, the FPR of the presented method is 10.8% better than k-means and 4% better than C-means methods. The FNR of proposed method is 6.13% superior to k-means and 8% superior to C-means algorithms. Moreover, the NPV of the implemented scheme is 9.14% better than k-means and 4.25% better than C-means methods. The FDR of the suggested scheme is 5% superior to k-means and 3.97% superior to C-means techniques. Similarly, the F1-score of proposed method is 12.5% better than k-means and 5.25% better than C-means techniques. Finally, the MCC-measure of the implemented scheme is 10.11% superior to k-means and 8.42% superior to C-means algorithms. Thus the enhancement of the proposed k-NN scheme has been validated successfully.

Similarly, the Hindi database was obtained from Table III, where, the suggested method in terms of accuracy is 10.7% better than k-means and 1.34% better than C-means algorithms. Likewise, the

proposed scheme concerning sensitivity is 11.29% superior to k-means and 6.11% superior to C-means techniques. Moreover, the specificity of the presented design is 8.47% better than k-means and 7.52% better than C-means techniques. In addition, the precision of proposed method is 9.5% superior to k-means and 11.49% superior to C-means techniques. Moreover, the FPR of the implemented system is 11.8% better than k-means and 9.76% better than C-means methods. The FNR of proposed method is 8.71% superior to k-means and 3.13% superior to C-means algorithms. Furthermore, the NPV of the suggested idea is 15.5% better than k-means and 13.52% better than C-means methods. The FDR of the proposed design is 1.82% superior to k-means and 0.67% superior to C-means techniques. Similarly, the F1-score of proposed technique is 14.6% better than k-means and 17.78% better than C-means algorithms. Finally, the MCC-measure of the implemented system is 5.26% superior to k-means and 4.21% superior to C-means schemes. Thus the improvement of the suggested k-NN method has been authenticated.

Moreover, the implemented system for benchmark database, concerning accuracy is 4.89% better than k-means and 5.85% better than C-means techniques. Correspondingly, the presented scheme about sensitivity is 11.29% superior to k-means and 6.11% superior to C-means schemes. Additionally, the specificity of the implemented design is 8.47% better than k-means and 7.52% better than C-means techniques. In addition, the precision of proposed scheme is 9.5% superior to k-means and 11.4% superior to C-means methods. Furthermore, the FPR of the implemented scheme is 11.84% better than k-means and 9.76% better than C-means techniques. The FNR of suggested means is 8.7% superior to k-means and 3.13% superior to C-means algorithms. Furthermore, the NPV of the implemented scheme is 15.57% better than k-means and 13.52% better than C-means methods. The FDR of the suggested scheme is 1.8% superior to k-means and 0.6% superior to C-means techniques. Similarly, the F1-score of proposed method is 14.6% better than k-means and 17.7% better than C-means techniques. At last, the MCC-measure of the implemented system is 5.2% superior to k-means and 4.2% superior to C-means algorithms. Thus the capability of the proposed k-NN model has been verified by the experimentations.

From Fig. 2, the overall accuracy of the suggested scheme for 10% learning percentage is 6.93% superior to k-means and 5.6% superior to C-means methods. Also, regarding learning percentage of 25%, the proposed method is 5.32% better than k-means and 2.05% better than C-means techniques. Also, the implemented method for learning percentage of 50% is 6.25% superior to k-means and 3.53% superior to C-means algorithms. In addition, learning percentage of 75% is 7.44% better than k-means and 3.86% better than C-means techniques. Finally, for 100% learning percentage, the proposed method is 7% superior to k-means and 3.82% superior to C-means techniques.

Table 2: *Performance of the proposed design for marathi database*

Methods	k-mean [21]	C-mean [22]	k-NN [20]
Accuracy	74.2	79. 9	82
Sensitivity	72.3	76.06	86.8
Specificity	75.1	74.6	88
Precision	73.2	78.32	81.89
FPR	82	77	74
FNR	79.6	81	75
NPV	75.5	79.56	83.1
FDR	83	82.14	79
F_1score	70	75.8	80
MCC	80	81.5	89

Table 3: *Performance of the proposed design for hindi database*

Methods	k-mean [21]	C-mean [22]	k-NN [20]
Accuracy	73.22	80. 9	82
Sensitivity	75.3	79.7	84.89
Specificity	77.8	78.6	85
Precision	79.09	77.35	87.4
FPR	85	83.42	76
FNR	83.6	79.31	76.9
NPV	72.76	74.52	86.18
FDR	82.15	81.22	80.68
F_1score	75.34	72.6	88.3
MCC	81	81.9	85.5

Table 4: *Performance of the proposed design for Benchmark database*

Methods	k-mean [21]	C-mean [22]	k-NN [20]
Accuracy	79.72	78. 91	83.82
Sensitivity	76.13	73.7	85.59
Specificity	71.36	80.02	84.23
Precision	73.90	76.3	86.24
FPR	86.73	89.65	76.3
FNR	83.5	80.31	77
NPV	78.3	79.34	83.5
FDR	87.1	80.45	84.36
F_1score	74.45	76.5	81.3
MCC	78.2	82.7	89.9

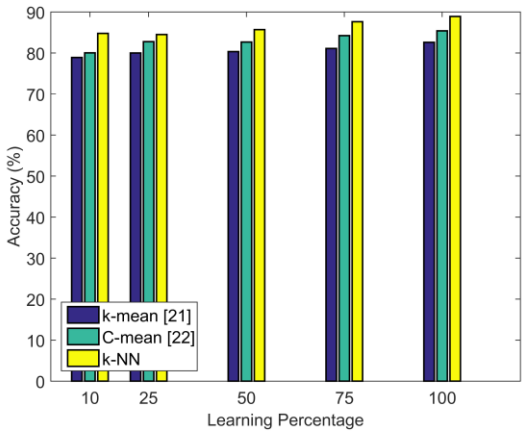

Fig. 2. *Overall accuracy of the proposed method.*

5. Conclusion

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As comma. This paper has presented SER using k-NN classifier for cepstral features. Moreover, the capability of the k-NN algorithm was validated using performance analysis and overall accuracy measures. From the experimentation, the proposed method for benchmark database, regarding accuracy is 4.89% better than k-means and 5.85% better than C-means techniques. Correspondingly, the presented scheme about sensitivity is 11.29% superior to k-means and 6.11% superior to C-means schemes. Additionally, the specificity of the implemented design is 8.47% better than k-means and 7.52% better than C-means techniques. In addition, the precision of proposed scheme is 9.5% superior to k-means and 11.4% superior to C-means methods. Furthermore, the FPR of the implemented scheme is 11.84% better than k-means and 9.76% better than C-means techniques. Thus the performance capability of the proposed model has been validated successfully.

References

[1] Shaoling Jing, Xia Mao, Lijiang Chen, "Prominence features: Effective emotional features for speech emotion recognition", Digital Signal Processing, vol. 72, pp. 216-231, January 2018.

[2] Qirong Mao, Guopeng Xu, Wentao Xue, Jianping Gou, Yongzhao Zhan, " Learning emotion-discriminative and domain-invariant features for domain adaptation in speech emotion recognition", Speech Communication, vol. 93, pp. 1-10, October 2017.

[3] Haytham M. Fayek, Margaret Lech, Lawrence Cavedon, "Evaluating deep learning architectures for Speech Emotion Recognition", Neural Networks, vol. 92, pp. 60-68, August 2017.

[4] Jiang Xiaoqing, Xia Kewen, Lin Yongliang, Bai Jianchuan, " Noisy speech emotion recognition using sample reconstruction and multiple-kernel learning", The Journal of China Universities of Posts and Telecommunications, vol. 24, no. 2, pp. 1-17, April 2017.

[5] Zhen-Tao Liu, Min Wu, Wei-Hua Cao, Jun-Wei Mao, Guan-Zheng Tan, "Speech emotion recognition based on feature selection and extreme learning machine decision tree", Neurocomputing, vol. 273, pp. 271-280, 17 January 2018.

[6] Sara Motamed, Saeed Setayeshi, Azam Rabiee, "Speech emotion recognition based on a modified brain emotional learning model", Biologically Inspired Cognitive Architectures, vol. 19, pp. 32-38, January 2017.
[7] Lijiang Chen, Xia Mao, Yuli Xue, Lee Lung Cheng, "Speech emotion recognition: Features and classification models", Digital Signal Processing, vol. 22, no. 6, pp. 1154-1160, December 2012.
[8] Yogesh C.K., M. Hariharan, Ruzelita Ngadiran, A.H. Adom, Kemal Polat, "Hybrid BBO_PSO and higher order spectral features for emotion and stress recognition from natural speech", Applied Soft Computing, vol. 56, pp. 217-232, July 2017.
[9] Moataz El Ayadi, Mohamed S. Kamel, Fakhri Karray, "Survey on speech emotion recognition: Features, classification schemes and databases", Pattern Recognition, vol. 44, no. 3, pp. 572-587, March 2011.
[10] Jae-Bok Kim, Jeong-Sik Park, "Multistage data selection-based unsupervised speaker adaptation for personalized speech emotion recognition", Engineering Applications of Artificial Intelligence, vol. 52, pp. 126-134, June 2016.
[11] Farah Chenchah, Zied Lachiri, "Speech Emotion Recognition in Acted and Spontaneous Context", Procedia Computer Science, vol. 39, pp. 139-145, 2014.
[12] Siqing Wu, Tiago H. Falk, Wai-Yip Chan, "Automatic speech emotion recognition using modulation spectral features", Speech Communication, vol. 53, no. 5, pp. 768-785, May–June 2011.
[13] Soroosh Mariooryad, Carlos Busso, "Compensating for speaker or lexical variabilities in speech for emotion recognition", Speech Communication, vol. 57, pp. 1-12, February 2014.
[14] Edmondo Trentin, Stefan Scherer, Friedhelm Schwenker, "Emotion recognition from speech signals via a probabilistic echo-state network", Pattern Recognition Letters, vol. 66, pp. 4-12, 15 November 2015.
[15] Leandro D. Vignolo, S.R. Mahadeva Prasanna, Samarendra Dandapat, H. Leonardo Rufiner, Diego H. Milone, "Feature optimisation for stress recognition in speech", Pattern Recognition Letters, vol. 84, pp. 1-7, 1 December 2016.
[16] Rahul B. Lanjewar, Swarup Mathurkar, Nilesh Patel, "Implementation and Comparison of Speech Emotion Recognition System Using Gaussian Mixture Model (GMM) and K- Nearest Neighbor (K-NN) Techniques", Procedia Computer Science, vol. 49, pp. 50-57, 2015.
[17] B. Yang, M. Lugger, "Emotion recognition from speech signals using new harmony features", Signal Processing, vol. 90, no. 5, pp. 1415-1423, May 2010.
[18] Khiet P. Truong, David A. van Leeuwen, Franciska M.G. de Jong, " Speech-based recognition of self-reported and observed emotion in a dimensional space", Speech Communication, vol. 54, no. 9, pp. 1049-1063, November 2012.
[19] Milton Sarria-Paja, Tiago H. Falk, "Fusion of auditory inspired amplitude modulation spectrum and cepstral features for whispered and normal speech speaker verification", Computer Speech & Language, vol. 45, pp. 437-456, September 2017.
[20] Zhenyun Deng, Xiaoshu Zhu, Debo Cheng, Ming Zong, Shichao Zhang, "Efficient kNN classification algorithm for big data", Neurocomputing, vol. 195, pp. 143-148, 26 June 2016.
[21] S. Borgwardt, A. Brieden, P. Gritzmann,"An LP-based k-means algorithm for balancing weighted point sets", European Journal of Operational Research, vol. 263, no. 2, pp. 349-355, 1 December 2017.
[22] Adrian Stetco, Xiao-Jun Zeng, John Keane, " Fuzzy C-means++: Fuzzy C-means with effective seeding initialization", Expert Systems with Applications, vol. 42, no. 21, pp. 7541-7548, 30 November 2015.