Paper:

Speech Emotion Recognition Based on PSO-SVR Using Personality Clusters

Abdul Rehman, Zhen-Tao Liu†, Min Wu, Wei-Hua Cao, Man Hao

School of Automation, China University of Geosciences, Wuhan 430074, China
Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
† Corresponding author, E-mail: liuzhentao@cug.edu.cn

Speech emotion recognition (SER) relies on training sets that are collected within a limited variety of speakers. As a result, their accuracy drops when tested with an alien dataset. We present a method that helps recognizing the personal similarities among similar speakers so that new speakers can be associated with a group of similar personalities. First, we group together the similar by personality speakers using K-means clustering. Then by using the PSO optimization, Support Vector Regression (SVR) model can be optimized to better differentiate the inter-group differences while learning intra-group characteristics. Emotional valance, as the target output of the regression model, is adjusted for each cluster accordingly. Experiments on the RAVDESS dataset, indicate the consistency of recognition accuracy when a new test subject is tested with a trained regression model thus highlighting the importance of the inclusion of dimensions of personality for SER.

Keywords: speech emotion recognition, personality analysis, optimization, affective computing

1. Introduction

Studying and understanding the emotions conveyed in the speech expressions is an important step toward designing dependable human-interactive machines. Machine learning methods that are usually employed in Speech Emotion Recognition (SER) require training under supervision using some labeled training data. Conventional approaches for SER generally assume that the feature distributions between training and testing set are identical [5, 12]. However, this assumption does not hold in many real scenarios (see [17]). In recent past, extensive efforts have been expended by researchers in the area of emotion recognition via speech by recognizing underlining common features of among different domains of speakers (e.g., [1, 3, 14]). Domains could be varying based on languages, gender, personality characteristics, and many other variables. One kind of variation lies in the domain of personality characteristics of speakers (such as big-five personality traits [7]), however, there can be many more different domains based on different kinds of variations. We focused our research on personality domain variation under the assumption that personality characteristics have similar underlying features as the emotional states and they both have overlapping features to some extent [26]. Basic emotions are momentary behaviors, whereas personality traits are expected to remain consistent over a long period of time. For continuous short-term estimation of emotions, long-term variables are vital to be considered [8]. Different methods have already been explored by many research works to work out these two convoluted dimensions (e.g., [16, 23]).

In SER, the valance-arousal framework is generally used to gauge the emotions on a continuous 2D plane [20]. This widely used emotional framework is still being questioned for their validity and effectiveness. In [25], the ordinal nature of emotions has been explored and arguments are presented in support of relative annotation as opposed to traditionally used absolute annotation. Many research works have proposed to use more than two affective dimensions to covers emotional space with more precision [6, 10, 24]. It is probable that the emotional frameworks on which SER systems measure up a speaker are prone to variations in the framework itself, which may lead to poor classification results because not all speakers fall into the hard defined boundaries of opposing emotions.

Various normalization techniques have already been experimented with, such as standardizing the min-max and zero-point for each speaker or for each domain [4]. Normalization leads to better performance across various domains so it is already being widely used in SER. But still, when a new speaker is tested on an already trained system, it doesn't work efficiently until it has collected enough data to learn normalization boundaries. To solve this cold-start problem, many adaptive solutions have been proposed recently [3, 14].

Our method is also aimed at solving the cold-start problem, by associating the new speaker to a group of speakers in the already trained dataset who have similar personalities. Hence the assumption that if two speakers have similar personalities, their emotions are will sound similar. We began this research with a hypothesis that different personality groups of speakers have different minima, maxima, and mean as compared to the global minima, maxima, and mean on the emotional valance scale. By controlling the variable of personality groups, we can minimize the variation in valance for each group's different emotions.

The rest of the paper is organized as follows. In Section 2, the proposed framework is explained which includes personality clustering, emotional valance regression, and emotional valance optimization for each personality cluster. Then in Section 3, experiments on the RAVDESS dataset are explained and then results are analyzed.

2. Proposed Framework

In an attempt to recognize differences while controlling the commonalities, we use Big-five personality traits and gender as the commonalities among a group of speakers while differentiating their basic emotions. Our method optimizes the valance value of different emotions to better suit their respective group of speakers who have similar personalities. It trains the valance regression model such that the target output valance is adjusted for each different group's each emotion. **Figure 1** shows a simplified flow diagram of our proposed framework. In essence, to better recognize the basic emotions, our model utilizes the estimated prediction of gender and personality classification to compensate for the variations among different groups of speakers. The overview of the approach is as follows. First, by using K-means clustering we group together the speakers who have been annotated with similar personality traits. Then we create an initial regression model using SVR with initial emotional labels as the target output. Then for each cluster, we use the particle swarm optimization (PSO) for optimizing the emotional framework. The optimized valance label $V_{c,e}$ for a certain cluster's certain emotion is explained in scalar terms as

$$V_{c,e} = V_{i,e} + R_{c,e} \quad \ldots\ldots\ldots\ldots \quad (1)$$

where c is the cluster index, e is the index of basic emotions (i.e., neutral, calm, happy, sad, angry, fearful, disgust, surprised), $R_{c,e}$ is the scalar difference of cluster c's emotion e from the initial target valance $V_{i,e}$ of emotion e. It should be noted that since each person is somewhat unique than everyone else, so there will always be differences among the individuals of a group. Individuals within the same group may have few similar traits but they might vary in other unaccounted domains or traits. Higher the similarity index of a group, higher will be the control of our personality controlling variable.

2.1. Similar Speakers Clustering

Our training model requires personality and gender labels (many other factors can be included depending on the availability of the annotated dataset). While gender is a dichotomous variable, personality varies on a spectrum of five dimensions (O.C.E.A.N) [7]; namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. There can be many clusters of speakers depending on the variations in annotation and parameters of clustering methods, therefore for some reduction in the classification depth is needed in order to make groups bigger.

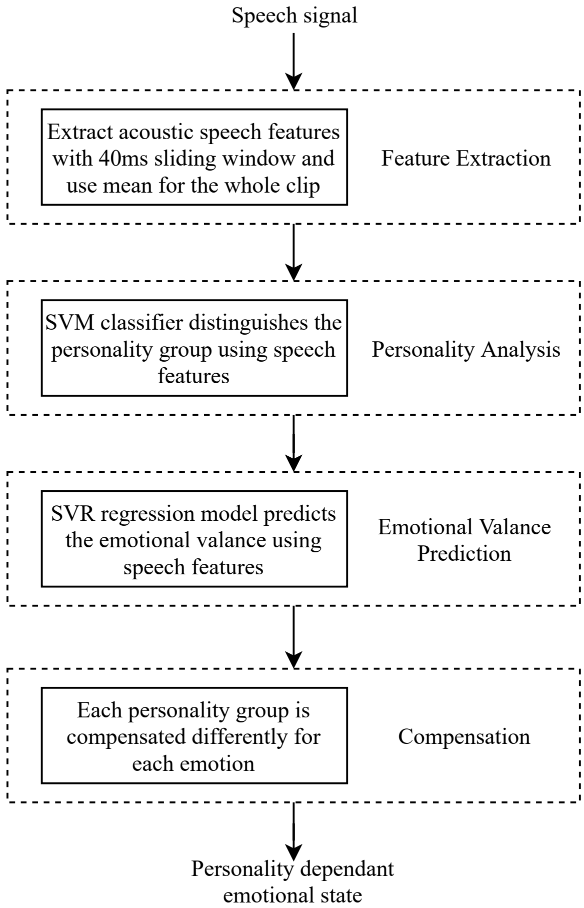

Fig. 1. Overview of the proposed framework. A Regression model is trained to estimate the valance label and a classifier model is trained to estimate cluster membership of the test subject. Using PSO, the regression model is optimized by searching each personality cluster's error minima. Final output label is predicted by compensating the difference between the optimal valance label for a certain cluster and the initial valance labels used for training.

We use an unsupervised clustering method that can figure out the similarly annotated speakers and group them together by tagging a cluster label. The clustering quality is dependant on the variation among speakers as well as the number or number (K) of clusters. It is advised that silhouette measure should be kept within acceptable limits. An iterative approach is used to group the data into a pre-determined K number of clusters by minimizing a cost function ζ that is calculated as

$$\zeta = \sum_{i=1}^{n} \left\| d_i - C_j \right\|^2 \quad \ldots\ldots\ldots\ldots \quad (2)$$

where C_j is the center of j^{th} cluster and is the center nearest to data object d_i, n is the number of elements in data set, which in our case is 6. Gender is a categorical variable, so for our purpose we can assume the two numeric

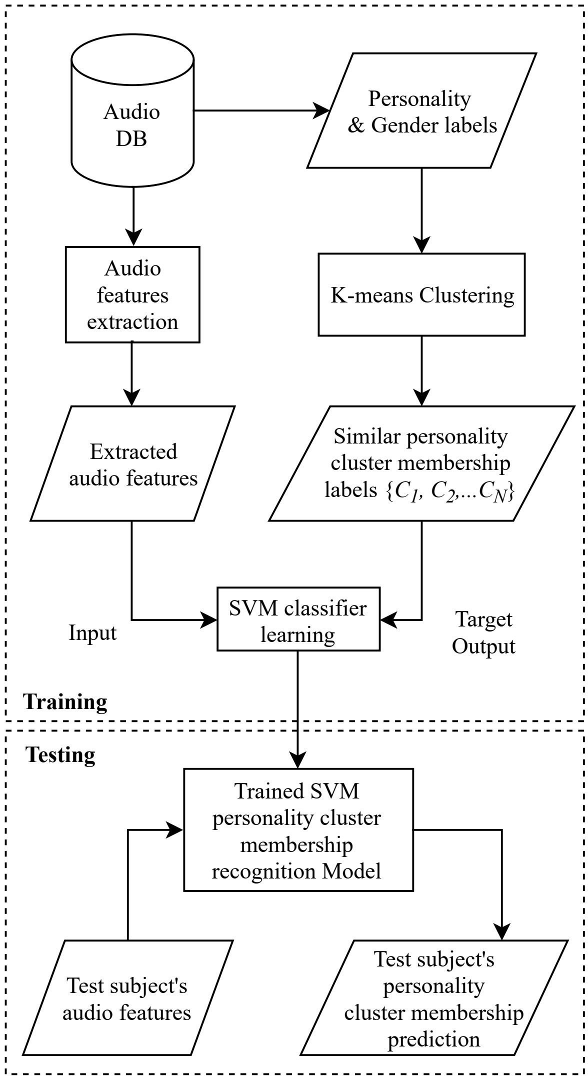

Fig. 2. Overview of unsupervised K-means clustering and SVM training for cluster member recognition.

values for two genders [2].

After unsupervised clusters are formed, speakers are automatically labeled by their assigned cluster indexes. Then in the next step, a linear SVM classifier is trained to recognize the cluster membership from extracted speech features. An overview of the clustering and training of personality label recognition is shown in **Fig. 2**.

2.2. Regression Model for Emotional Valance

In affective computing community, emotions are mostly measured on two dimensions of arousal-valance scale. In this research work, we only focused on the valance dimension due to the unavailability of a dataset with a high variance in arousal. **Figure 3** shows the regression training model and the optimization process. We use the extracted acoustic speech features as the input to estimate value of the emotional valance. This regression model is trained repeatedly using marginally adjusted target valance outputs until the *RMSE* (Root Mean Square Error) minima is reached. The regression model can be based on various intelligent algorithms. SVR is one of the frequently used regression models for prediction of the emotional valance (such as by [9]). An analysis by [11] also regards it as a high-performance model. We use the extracted speech features as input and the emotional valance as the target output of a liner kernel SVR model. Using the K-folds validation method, the validation prediction of each training subject is measured then *RMSE* is calculated for the trained model. Lower the *RMSE*, the better the model is considered. Our objective is to minimize the *RMSE* for the whole training set, so we adjust the target valance output for each personality group and train the model again until *RMSE* minima are reached.

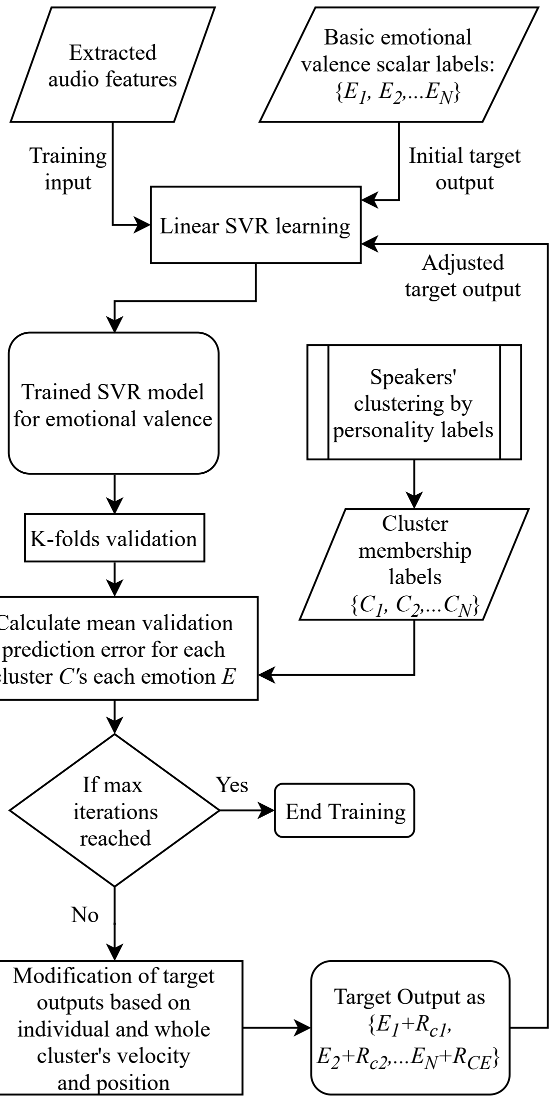

Fig. 3. Proposed method using the Particle Swarm Optimization (PSO) for searching target outputs with least error minima for each cluster. Each cluster is treated as a swarm with its members as individual particles. After minima are reached, the optimized target output value of the whole cluster is averaged into one to be used for testing.

2.3. PSO Optimization of the Target Outputs

For the optimization of the regression model, each cluster is treated as a swarm and its members as particles. Our goal is to locate the target output set (that can be different for each cluster) that minimizes the validation of the regression model for the whole training set. When all cluster swarms narrow down on their optimal target outputs, the whole training set achieves better *RMSE* and regression R^2. Initially, all clusters have the same target outputs, but with each iteration, clusters move away from the initial targets to their own best performance targets. After optimization, the target outputs of all members of a cluster are averaged into a cluster average. In each iteration, the velocity and position of the target output of each member of the cluster are updated by using the combination of both, the individual best and the cluster's best minima. The change in target in next iteration can be defined by velocity $v_i(t+1)$ as

$$v_x(t) = \widehat{x_i}(t) - x_i(t) \quad \quad (3)$$

$$v_c(t) = c(t) - x_i(t) \quad \quad (4)$$

$$v_i(t+1) = wv_i(t) + c_1 r_1[v_x(t)] + c_2 r_2[v_c(t)] \quad \quad (5)$$

and the position x_i is updated in next iteration as

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad \quad (6)$$

where $x_i(t)$ is the current position of the individual i for iteration t, $\widehat{x_i}(t)$ is the individual best minima, $v_x(t)$ is individual's displacement from its own best minima, $v_c(t)$ is the displacement from cluster's best minima $c(t)$. c_1 and c_2 are acceleration coefficients which keep the balance between the convergence speed and explored area. And r_1, r_2 are the random values $(0 \leq r_1, r_2 \leq 1)$ to promote search in the unexplored space.

When a test subject is tested with the optimized trained regression model, it will output the valance label, and the personality cluster classifier model will output the cluster membership labels. Using a compensating look-up table, both outputs are used to figure out the valance label in the initial framework terms.

3. Experimentation

To test our presented approach, the experiments are carried out in python 3.7.0 and MATLAB R2018a on the computer of the 64-bit Windows 10 system with 8G memory. CPU of the computer used in experiments is Intel Core i7-8550U.

3.1. RAVDESS Dataset

The dataset we used for experimentation is Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [13]. This dataset was specifically designed for research on basic emotions and originally does not include the Big-five personality annotations. But due to the quality, clarity, uniformness of the recordings, we decided to use it by first annotating the personality factors using the same method as used by Big-five annotated dataset collectors [18, 22]. This dataset comprises of recordings by 23 gender-balanced (12/11) actors performing the act of various basic emotions. During annotation collection, we only used the emotionally-neutral clips so that there is no bias because of short-term emotions. Out of total of 2452 recordings available, 2392 recordings of 23 actors were used for feature extraction and classification learning.

Table 1. Description of dataset used for experimentation.

Attribute	Description	Scale
Gender	12 M, 11 F	Binary (1 or 2)
Valence [13]	Neutral, calm, happy, sad, angry, fearful disgust, surprised	Initial valence is assumed 1,2,...,8 respectively
Personality	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism	Each trait rated by 4 raters from -4 to +4 using BFI-10 [19]
Speech features	41 acoustic features extracted using YAAFE [15]	Normalized

3.2. Speech Features Extraction

There are few open source tools available for audio feature extraction and are being used for SER research works [21]. In our work, we chose to use YAAFE (Yet Another Audio Feature Extraction) toolbox [15] v0.64 to extract features from the speech database. Total of 41 features were extracted from each recording for analysis. Each audio WAV file is segmented into 1024-bytes blocks consecutive blocks overlapped by 50 percent. Then each block is used for feature extraction separately. Average for the whole audio clip for each feature is then compiled from feature values of all blocks.

3.3. K-means Clustering and Classifier Learning

RAVDESS database originally does not include personality annotations, so we annotated each speaker ourselves using the BFI model [18]. All speakers were rated for their personality traits on a scale from -4 to +4 by 4 raters. Then a rounded average of all raters was assumed as the apparent scores of five traits for all speakers. Six input variables, i.e., 5 personality trait scores, and gender were used to perform K-means clustering which led to 7 clusters of all 23 speakers; each cluster containing 3 or 4 speakers. Training/testing sets were divided by 20/3, taking 3 test subjects from 3 different clusters. RAVDESS dataset has relatively fewer speakers but the amount of data per speaker is relatively high; around 100 recordings per speaker, varying across 8 basic emotions. For 7 clusters, a mean Silhouette measure of 0.554 was achieved.

The centers of clusters on 5 personality traits (on a scale of -4 to +4) and gender are shown in **Table 2**.

Extracted features were then fed into a machine learning phase where classifiers were trained to classify between the cluster memberships. The SVM binary classification algorithm searches for the best hyperplane that splits the feature data into two classes. MATLAB Statistic and Machine Learning Toolbox was used for both types of classification. Liner and quadratic kernel SVM were both tested to see which one yields better results. K-fold Cross Validation method with 5 folds was used for validation of the training set.

Table 2. Centers of 7 clusters on 6 personality dimensions and number of speakers (N_c) in each cluster.

Traits	Cluster No.						
	C1	C2	C3	C4	C5	C6	C7
O	-2	4	0	1	-3	-4	2
C	3	-1	3	2	-1	2	-2
E	2	4	2	-3	-4	-4	-2
A	-1	3	2	2	0	-3	2
N	-1	2	2	3	-4	-2	-2
M/F	1	0	1	0	1	1	1
N_c	3	4	4	3	3	4	2

3.4. Regression Model and Optimization

Our dataset is labeled with 8 different basic emotions labels as scalar values from 1 to 8. This scalar model is assumed based on the valance of emotions in a non-linear order of neutral, calm, happy, sad, angry, fearful, disgust, surprised. Since the activation/arousal dimension is not used in our model, the non-linearity of valance helps the regression training model to learn the complexity. Initial labels were fed into the optimization scheme as target outputs. A hundred iterations of PSO were tested, and it leads to different target outputs for different emotions for different clusters. Finally, a cross-referencing table was created that refers to the optimized valence for each cluster's each emotion in order to compensate the estimated value of valance to predict the initial value of valance. **Table 3** shows the final centers for each cluster's each emotion. For testing, three test speakers (312 recordings out of 2392) were tested with the trained model. First, their personality cluster was predicted, then the valance value for each emotion was predicted, then by cross-referencing the compensation table, the emotion with the nearest predicted valance is stated as output.

3.5. Experimental Result and Analysis

The results of the experiment on the RAVDESS dataset for personality cluster membership prediction are shown in **Table 4**. For cluster membership prediction a quadratic kernel SVM performs with better accuracy as compared to the linear kernel. So for personality based optimization of PSO-SVR, only the quadratic kernel SVM for cluster membership prediction was used.

Table 3. Initial Valance (IV) and the final PSO optimized valance centers for each cluster's each emotion.

IV	Optimized valance centers for clusters						
	C1	C2	C3	C4	C5	C6	C7
1	-1.4	-1.2	-1.2	-1.5	-1.4	-1.6	-0.7
2	0.72	1.54	1.07	0.82	0.84	0.47	1.80
3	2.08	1.54	1.83	1.95	1.51	1.72	2.28
4	4.09	4.70	4.33	4.21	4.27	3.92	4.63
5	4.87	5.08	5.29	5.24	5.06	5.16	5.25
6	7.09	7.40	7.35	6.85	7.08	7.25	7.76
7	8.94	8.96	8.78	9.24	9.13	9.48	9.68
8	10.8	10.4	10.8	10.2	10.4	10.5	10.4

A comparison of results of an SVR model trained using the initial valance values vs with the optimized valance values for each cluster is given in **Table 5**. Coefficient of determination R^2 and $RMSE$ improves when personality clustering is used in the optimization scheme. This could be due to the broader definition of variability in the annotations. One big observable difference is the difference between the training and testing set regression accuracy; without the dedicated optimization, R^2 drops by 24% for the testing set. This clearly indicates the importance of similar speakers' group association that learns to distinguish between the 7 clusters. We divided the training and testing set such that test speakers were not used in training set, but 2 or 3 personality-wise similar speakers were still in the training set. When the SVM classifier distinguishes the new member as similar to a certain cluster just by using the audio features, it narrows down the error margin for the SVR. Hence, adding a personality recognition feature is narrowing down on the variation in the apparent valance of a certain emotion. In $RMSE$, the drop is not significant, reason might be the non-linear valance labeling of emotions. In the future, we intend to experiment with a dataset with a more linear valence labeling as well as the arousal labeling. Meanwhile, the R^2 values are a prominent indicator that optimizing the valance scale using personality tuned framework increases the reliability of an SER system when new test subjects are tested on it.

Table 4. Validation accuracy of SVM classifier trained for personality cluster membership prediction using 41 audio features as input. Quadratic kernel achieves better accuracy as compared to the linear kernel.

Model	Parameters	Accuracy
SVM	Linear Kernel	66.3%
SVM	Quadratic Kernel	83.7%

4. Conclusion

Speech emotion recognition is currently facing a cold-start challenge, i.e., the recognition accuracy drops as soon as a new speaker is tested with an already trained

Table 5. Results of the experiment on RAVDESS dataset for comparison between a linear SVR model trained with the global initial values of emotional valance vs the linear SVR model that uses the PSO optimized values of emotional valances based on personality cluster membership prediction to minimize the *RMSE*.

Target Output Valence	*RMSE*		R^2	
	Train	Test	Train	Test
Global Initial Values	1.56	5.58	0.32	0.03
Clusters' PSO Values	1.53	3.02	0.26	0.24

system. The variation in speakers causes the recognition accuracy to steep lower when training and testing sets are unmatched by a number of factors. To keep the accuracy from dropping across varying speakers, our approach is to differentiate the scale on which emotions are measured for different groups. We presented a method that utilizes the personality of speakers to classify them into similar groups so that short-term emotions can be better recognized by a supervised regression model. By first clustering the annotations of personality using the K-means algorithm we proposed to train a classifier to classify similar speakers, then any new speaker is automatically associated with its nearest similar group of speakers. This cluster association helps our system in recognizing common paralingual features of similar speakers. By using the PSO optimization, our method optimizes the better suited valance value for each group's each emotion instead of using the global valance values. With experiments on the RAVDESS dataset, results indicate an improvement in the reliability of SER using our method.

With the upcoming challenges in SER, there is a higher need for speaker adversarial methods, thus keeping the tracks of variations in speakers has become important. We showed that with only within a sample of 23 speakers, the gap between the trained and test accuracy reduces when some long-term characteristics of speakers are considered. However, using this method on a vast scale requires a high labeling cost. In the future, further research needs to be conducted to make this process easier, simpler and less supervised.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants 61403422 and 61273102, the Hubei Provincial Natural Science Foundation of China under Grant 2018CFB447 and 2015CFA010, the Wuhan Science and Technology Project under Grant 2017010201010133, the 111 project under Grant B17040, and the Fundamental Research Funds for National University, China University of Geosciences (Wuhan) under Grant 1810491T07.

References:

[1] M. Abdelwahab and C. Busso, "Domain adversarial for acoustic emotion recognition", IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(12):2423–2435, 2018.

[2] A. Ahmad and L. Dey, "A k-mean clustering algorithm for mixed numeric and categorical data", Data & Knowledge Engineering, 63(2):503–527, 2007.

[3] J. Bang, T. Hur, D. Kim, J. Lee, Y. Han, O. Banos, J.-I. Kim, S. Lee, et al., "Adaptive Data Boosting Technique for Robust Personalized Speech Emotion in Emotionally-Imbalanced Small-Sample Environments", Sensors, 18(11):3744, 2018.

[4] K. Brady, Y. Gwon, P. Khorrami, E. Godoy, W. Campbell, C. Dagli, and T. S. Huang, "Multi-modal audio, video and physiological sensor learning for continuous emotion prediction", In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, pp. 97–104. ACM, 2016.

[5] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases", Pattern Recognition, 44(3):572–587, 2011.

[6] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional", Psychological science, 18(12):1050–1057, 2007.

[7] L. R. Goldberg, "An Alternative "Description of Personality": The Big-Five Factor Structure", Personality and Personality Disorders: The Science of Mental Health, 7:34, 2013.

[8] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions", Image and Vision Computing, 31(2):120–136, 2013.

[9] J. Han, Z. Zhang, F. Ringeval, and B. Schuller, "Prediction-based learning for continuous emotion recognition in speech", In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 5005–5009. IEEE, 2017.

[10] S. S. Jang and Y. Namkung, "Perceived quality, emotions, and behavioral intentions: Application of an extended Mehrabian–Russell model to restaurants", Journal of Business Research, 62(4):451–460, 2009.

[11] Z. A. Khalifelu and F. S. Gharehchopogh, "Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation", Procedia Technology, 1:65–71, 2012.

[12] N. Khan, "A Review on Speech Emotion Recognition", IJOSTHE, 3(2):6–6, 2016.

[13] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English", PloS one, 13(5):e0196391, 2018.

[14] K. Mannepalli, P. N. Sastry, and M. Suman, "A novel Adaptive Fractional Deep Belief Networks for speaker emotion recognition", Alexandria Engineering Journal, 56(4):485–497, 2017.

[15] B. Mathieu, S. Essid, T. Fillon, J. Prado, and G. Richard, "YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software.", In ISMIR, pp. 441–446, 2010.

[16] M. A. Nicolaou, H. Gunes, and M. Pantic, "Output-associative rvm regression for dimensional and continuous emotion prediction", Image and Vision Computing, 30(3):186–196, 2012.

[17] O. E. N. Noi, M. Qirong, G. Xu, and Y. Xue, "Coupled Unsupervised Deep Convolutional Domain Adaptation for Speech Emotion Recognition", In 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), pp. 1–5. IEEE, 2018.

[18] B. Rammstedt, "The 10-item big five inventory", European Journal of Psychological Assessment, 23(3):193–201, 2007.

[19] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German", Journal of research in Personality, 41(1):203–212, 2007.

[20] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant.", Journal of personality and social psychology, 76(5):805, 1999.

[21] M. Swain, A. Routray, and P. Kabisatpathy, "Databases, features and classifiers for speech emotion recognition: a review", International Journal of Speech Technology, 21(1):93–120, 2018.

[22] A. Vinciarelli and G. Mohammadi, "A survey of personality computing", IEEE Transactions on Affective Computing, 5(3):273–291, 2014.

[23] M. Wollmer, B. Schuller, F. Eyben, and G. Rigoll, "Combining long short-term memory and dynamic bayesian networks for incremental emotion-sensitive artificial listening", IEEE Journal of Selected Topics in Signal Processing, 4(5):867–881, 2010.

[24] W. Wundt, "Elem Folk Psyc: Esc", volume 7, Routledge, 2013.

[25] G. N. Yannakakis, R. Cowie, and C. Busso, "The ordinal nature of emotions", In 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 248–255. IEEE, 2017.

[26] C.-W. Yu and J.-Y. Choi, "Behavior decision model based on emotion and dynamic personality", International Conference on Control, Robotics and Systems, pp. 101–106, 2005.