Proceedings Paper:
Gong, X, Yuan, R, Qian, H et al. (2 more authors) (2021) Emotion Regulation Music Recommendation Based on Feature Selection. In: New Trends in Intelligent Software Methodologies, Tools and Techniques. The 20th International Conference on Intelligent Software Methodologies, Tools, and Techniques (SoMeT 2021), 21-23 Sep 2021, Cancun, Mexico/Virtual. IOS Press , pp. 486-495. ISBN 978-1-64368-194-8

https://doi.org/10.3233/faia210047

Emotion Regulation Music Recommendation based on Feature Selection

Xiaoliang Gong[a] Ruiyi Yuan[a] Hui Qian[a] Yufei Chen [a,1] Anthony G Cohn [a, b]
[a] *School of Electronics and Information Engineering, University of Tongji, Shanghai, China*
[b] *School of Computing, University of Leeds, Leeds, UK*

Abstract. Chinese traditional music has been proved to be effective in emotion regulation for thousands of years. Five different groups of Chinese traditional music which have been proved can regulate different emotions (Angry, Depressed, Feverish, Desperate, Sorrowful) in the literature. 54 audios features are extracted by using the Librosa library for each music group. Five features are manually selected using histogram analysis which show significant difference between the five groups of music. Combined with KNN, SVM and Deep forest classification algorithms, the five manually selected audio features are shown to have better classification performance than traditional feature selection algorithms, like PCA and LDA. We hypothesize that these five significant audio features may be the underlying basis why so such music can effectively perform emotion regulation. Based on this classification models, prototype emotion regulation music recommendation interface (TJ-ERMR) was built that can be used for music therapy. In the future, we will use this classification model to find more music to expand the initial repertoire of our music recommendation system.

Keywords. Emotion Regulation, Music Recommendation, Librosa, Feature selection, Deep Forest

1. Introduction

Nowadays, more and more people live under high pressure. Long-term depression, eventually leading to increased levels of depression in the population is increasing year by year. Music therapy has a long history and tradition in China. Chinese ancestors have recognized the relationship between music, medicine and treatment[1, 2]. In fact, the use of music as an emotion regulation strategy is wide spread across the world[3]. Receiving music therapy means that an individual person listens to different kinds of music under the music therapists' guide. Music therapy varies according to the purpose of treatment, the physiological and psychological conditions of the individual, and the actual treatment environment to make custom therapeutic repertoires. It is expensive and difficult to find a music therapist. The reality is that there are a lot of people who need emotional regulation, but music therapists are lacking. An automated therapist may thus have utility. Moreover, listening to music at home is a popular way to decompress one's negative emotions.

Emotional expression forms are diverse. According to the intensity, persistence and tension of occurrent emotion, it can be divided into 3 levels. Previous work to estimate

[1] Corresponding Author, Yufei Chen, College of Electronic and Information Engineering, Tongji University, Shanghai, China; E-mail: yufeichen@tongji.edu.cn.

a person's instantaneous emotions has been based on input data such as user interaction data with the keyboard and mouse [4], or speech emotion recognition[5]. This has then been combined with machine learning algorithms to estimate instantaneous emotion and then recommend music to the user. When people are in a negative mood for several days, his/her mental state tends to be morbid. If the music therapy carries in time to regulate the negative emotion, then they will experience less pain. To the authors' knowledge, most previous studies have focused on the user's instantaneous emotion estimation rather than the selection of appropriate music that can do the music therapy. Such personalized music recommendation systems make recommendations according to the user's personal instantaneous preferences, but do not apply adjustments based on the characteristics of music. According to the Chinese traditional medical records, Xu al. explored a composition of Chinese traditional music therapy repertoire. Under the guidance of modern science and technology, the musical efficacy of these tracks is observed and demonstrated [6].

In this work, we use feature engineering method based on Librosa[7] to extract 54 audio features from five groups of Chinese traditional music which have been proved to be effective in the regulation of five different emotion(Angry, Depressed, Feverish, Desperate, Sorrowful) [1,2]. We then used feature selection and classification algorithms to find the significant audio features of emotion regulation music. These features of Chinese traditional music may be the cause of the effectively emotional regulation. Finally, the authors built a protype of an emotional regulation music recommendation user interface with an expanded Chinese traditional music therapy repertoire based on the feature selection results achieved in this work.

2. Methodology

2.1. Dataset

The Chinese traditional music dataset is collected from two papers that showed that music can regulate emotion [1,2]. After collecting and classifying all the songs from the two documents, we build a five-class music data set, which has 79 songs in total. The number of each type of music is shown in Table 1.

Table 1. The size of the Emotion Regulation Music class dataset

Emotion	**Angry**	**Depressed**	**Feverish**	**Desperate**	**Sorrowful**
# pieces	13	20	18	15	13

2.1.1. Feature extraction

Librosa[7] is a python package for music and audio analysis which we use for feature extraction. It provides building blocks suitable to create music information retrieval systems.

- Rhythm features

  Tempo

Tempo is the speed or pace of a given piece. In classical music, tempo is typically indicated with an instruction at the start of a piece (often using conventional Italian terms) and is usually measured in beats per minute (or bpm).

- Spectral features

  Spectral representations—the distributions of energy over a set of frequencies—form the basis of many analysis techniques in MIR and digital signal processing in general. A variety of spectral representations are implemented in the Librosa feature module. Some main features in music classification are as follows:

  MFCC

  In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC.

  Spectral centroid

  The spectral centroid is a measure used in digital signal processing to characterize a spectrum. It indicates where the center of mass of the spectrum is located. Perceptually, it has a robust connection with the impression of brightness of a sound.

  Zero-crossing-rate

  The zero-crossing rate (ZCR) is the rate at which a signal changes from positive to zero to negative or from negative to zero to positive[8]. Its value has been widely used in both speech recognition and music information retrieval, being a key feature to classify percussive sounds[9].

  ZCR is defined formally as:

$$\mathrm{ZCR} = \frac{\sum_{t=1}^{T=1} 1_{R<0}}{\mathrm{T}-1} (S_t S_{t-1}) \tag{1}$$

  where s is a signal of length T and $1_{R<0}$ is an indicator function.

  Chroma

  Chroma in western music closely relates to the twelve different pitch classes. Chroma-based features are a powerful tool for analyzing music whose pitches can be meaningfully categorized (often into twelve categories) and whose tuning approximates to an equal-tempered scale. The underlying observation is that humans perceive two musical pitches as similar in color if they differ by an octave. Based on this observation, a pitch can be separated into two components, which are referred to as tone height and chroma[10]. One main property of chroma features is that they capture harmonic and melodic characteristics of music, while being robust to changes in timbre and instrumentation.

2.1.2. Feature visualization

A kernel density estimate (KDE) plot is a method for visualizing the distribution of observations in a dataset, which is analogous to a histogram. KDE plot represents the data using a continuous probability density curve in one or more dimensions[11].

2.1.3. Data Normalization

In terms of feature rescaling, we use the common method min-max normalization, which also known as min-max scaling or min-max normalization, is the simplest method and consists in rescaling the range of features to scale the range in [0, 1] or [−1, 1]. The general formula for a min-max of [0, 1] is given as:

$$X^* = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

2.2. Feature Selection

In order to reduce the time and space complexity of the model and improve the robustness of the model on small data sets, researchers usually reduce the dimension of feature data[12]. In this work, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) are applied to do the feature selection, which represent supervised and unsupervised dimensionality reduction methods respectively. Meanwhile, to comprehend the features better, a manual feature strategy is also used.

- PCA

  PCA is a mathematical algorithm which is used to reduce the dimensionality of the data while keep most of the variation in the data set[13]. We experiment with different values of the main components, and keep five main components.
- LDA

  LDA is an algorithm to reduce the dimensionality of the data by finding the projection hyperplane that minimizes the interclass variance and maximizes the distance between the projected means of the classes[14].

  Our work needs a projection from multi class to a low dimensional space; the low dimensional space projected to is not a straight line, but a hyperplane. The dimension of the low dimensional space we project to is d and the value are between 2 and n – 1 (n is the number of the classes).
- Manually Selected Features

  We aim to give an explanation of the most significant features distinguishing between the five groups of emotion regulation music. We used histograms to visualize all the 54 features [22] and then manually selected the 5 most different features among the five groups of music, shown in Figure 1.

2.3. Classification

In order to compare the characteristics of different music classification algorithms, three music classification methods are chosen to classify the music dataset. The following are the advantages of each algorithm: KNN algorithm requires the less computing resources and is easy to implement; SVM is suitable for data sets with small samples of highlatitude features; Deep-Forest algorithm is novel and has fewer hyperparameters and faster convergence speed.

- K-nearest neighbor (KNN)

  The KNN classification algorithm is a theoretically mature method and one of the most common machine learning algorithms. It has low time complexity, and has no assumptions about data and is insensitive to outliers, so it is often used for comparison with SVM[15]. In this work, the KNN parameter k is optimized from 1 to 50 by cross-validated grid-search over a parameter grid.

- Support vector machine (SVM)

  A SVM is a machine learning method based on statistical learning theory developed in the 1990s. Good nonlinear processing ability and excellent generalization ability enable it to deal with high-dimensional small sample problems well, and avoid the local optimal problem in traditional neural network[16]. It is a classification model commonly used in audio signals classification[17, 18].

  Using SVM classification, two parameters need to be determined: the penalty coefficient C and the kernel function. The larger C is, the greater the penalty for the error-separating samples, so the accuracy is higher in the training samples, but the generalization ability is reduced.

  In this paper, the specific value of parameter penalty coefficient C is obtained by optimization. The penalty coefficient C was optimized over the values 1, 10, 100, 1000, 10000 and kernel function was chosen as a linear kernel function by default.

- Deep Forest (DF)

  Deep forest, also known as multi-Grained Cascade Forest (gcForest), is a decision tree-based learning method, which integrates and connects a forest composed of decision trees.[19].

  Because the deep forest model has low computational overhead, fewer super parameters, good adaptability for different sizes of data sets, especially small data sets, and easier theoretical analysis, it has been used in signal and image classification in recent years[20, 21].

  In this work, to reduce the complexity of the model, we manually adjust the parameters n_estimators which control the number of estimators in each cascade layer. n_trees which decide the number of trees in each estimator and max_depth which limit the maximum depth of tree. And n_estimators, n_trees and max_depth are parameters names in the deep forest[19].

2.4. Parameter optimization & Cross validation

In the process of classification, we use nested cross validation[22] to optimize the super parameters of the classifier and verify the model. The specific steps are as follows:

Step1: A 10-fold cross validation is carried out on the whole dataset, which is the outer layer of nesting.

Step2: In each outer cross validation there is a training dataset and a test dataset. And another 10-fold cross validation on the classifier based on the specific parameter which is the inner layer of nesting. The average accuracy of inner cross validation was used as the criterion to select the best model parameters.

Step3: The 10-fold outer cross validation results are considered as the model result.

2.5. Evaluation

Because it is a multi-classification task, in addition to accuracy, we also calculate F1-micro, F1-macro and F1-weighted to evaluate the model[23].

2.6. Computing resource

The classification methods used in this work is implemented in the python package “sklearn” and “deepforest”. All the work in this study is calculated on Windows10 with CPU 2.80GHz, RAM8.0GB.

2.7. User Interface

PyQt5 package is mainly used to implement the user interface. A music visualization function is added to make user feel more immersively by using the FuncAnimation package from matplotlib. The current system prototype has eight functions: selecting the current mood, loading therapy music library, playing therapy music, switching interface language, displaying playback time and progress, displaying song name, visualizing music and record user satisfaction feedback.

3. Results and Discussion

This work is an initial work to build a music recommendation system that can be used for self-help music therapy. In this work, the authors use machine learning to explore the characteristics of Chinese traditional music tracks recorded in the literature that can be used for music therapy. Next step, the authors wills aim to expand the music library for the self-developed emotion regulation music recommendation system. The ultimate goal is to create a self-help music therapy application software to provide timely professional music therapy services.

3.1. Manually Selected Five Significant Features of Five Group Emotion Regulation Music

We used Librosa to extract 54 standard audio features which includes most of the spectral features and rhythm feature. There are three kinds of features: mean value, standard deviation, and variance of most features, except the *tempo*, *total_beats* and *average_beats*, as shown in Figure 1(a). As Table 2 illustrates, the classification performance of different audio features combinations with different classification algorithms are regular. The combination of 54 features shows excellent classification effect on SVM and DF classifiers.

To reduce computation costs and enhance the explanation of input features, three kinds of feature selection methods are utilized. We found that feature selection improves the classification performance, especially the four features selected by LDA (a reduction of 92.26%, but still with good performance). With this combination of four features the classification results are changed about + 43.39%, -1.25%, -5%, respectively using KNN, SVM and DF as classifiers.

Compared with the four features selected by LDA, the five manually selected features perform better. In Figure 1(b), it can be clearly observed that these five features (*average beats, mfcc mean, rmse mean, cent mean and contrast mean) display good separation across the five classes.* There are obvious differences, like the half-height width of distribution, the maximum peak height etc. The classification performance of five manually selected features improves classification accuracy by 50.89% and 8.75%, compared to the KNN and SVM classifiers. A more detailed analysis of the difference among the audio features of each emotion regulation music will be conducted in the future.

3.2. Emotion Regulation Music Recommendation based on Feature Selection

The previous works[4,5] analyzed the listener's current emotional state by analysis of the operation of the keyboard and mouse. Their music recommendations focus on the user's personal favorite which may not suitable for the emotion regulation at that moment. Based on our research results, we implemented a prototype user interface to an emotion regulation music recommendation platform (TJ-ERMR). Using this TJ-ERMR platform, people can positively regulate their emotions when they are experiencing unpleasant emotions. As shown in Figure 2, users can select their current emotional state in the upper left corner. A user clicks the ‘OK’ button to load the music list corresponding to mood and then clicks the play button to listen to music and receive music treatment. It can be used for daily music therapy in the future.

However, there are few known tracks available for music therapy. At present, the authors have collected 79 tracks across the five categories. By using the classification models developed in this paper, the number of tracks in each category can be greatly increased. This work is a contribution towards the use of artificial intelligence algorithms to explore more music tracks suitable for specific emotional regulation. Of course, for actual deployment experts in music and medicine should also be involved to validate the results and the methodology. In the future, it will expand this ERMR repertoire of music.

a)

Totally 54 Audio features		
tempo	total_beats	**average_beats**
chroma_stft_mean	chroma_stft_std	chroma_stft_var
chroma_cq_mean	chroma_cq_std	chroma_cq_var
chroma_cens_mean	chroma_cens_std	chroma_cens_var
melspectrogram_mean	melspectrogram_std	melspectrogram_var
mfcc_mean	mfcc_std	mfcc_var
mfcc_delta_mean	mfcc_delta_std	mfcc_delta_var
rmse_mean	rmse_std	rmse_var
cent_mean	cent_std	cent_var
spec_bw_mean	spec_bw_std	spec_bw_var
contrast_mean	contrast_std	contrast_var
rolloff_mean	rolloff_std	rolloff_var
poly_mean	poly_std	poly_var
tonnetz_mean	tonnetz_std	tonnetz_var
zcr_mean	zcr_std	zcr_var
harm_mean	harm_std	harm_var
perc_mean	perc_std	perc_var
frame_mean	frame_std	frame_var

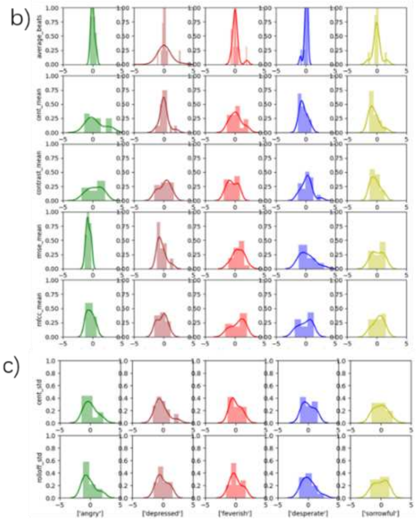

Figure 1. a) 54 audio features used in this work, and the five manually selected significant features in bold, b) KDE plots of the five most different audio features among five groups of emotion regulation music, c) By contrast, two features which have almost the same values across all five groups of music.

Table 2. The classification performance of different audio features combinations with different classifier.

Methods \ Classification of 5 categories	Accuracy(%)	F1-micro(%)	F1-macro(%)	F1-weighted(%)
KNN(54)	46.61±2.06	46.66±2.06	35.00±2.25	40.03±1.92
PCA(5)+KNN	49.64±2.33	49.64±2.33	38.93±2.66	41.99±2.41
LDA(4)+KNN	90.00±0.87	90.00±0.87	84.80±2.42	87.67±1.39
PCA(5)+LDA(4)+KNN	88.75±1.70	88.75±1.70	85.53±2.89	86.50±2.48
Manual(5)+KNN	97.47±0.25	97.47±0.25	97.33±0.57	97.46±0.39
SVM(54)	90.00±1.19	90.00±1.19	90.73±1.12	90.08±1.00
PCA(5)+SVM	72.32±2.07	72.32±2.07	67.20±1.64	67.50±2.68
LDA(4)+SVM	88.75±1.39	88.75±1.39	84.87±2.82	87.33±1.81
PCA(5)+LDA(4)+SVM	81.07±1.93	81.07±1.93	75.47±2.89	77.22±2.73
Manual(5)+SVM	98.75±0.14	98.75±0.14	97.60±1.12	98.75±1.05
Deep-Forest(54)	96.25±0.64	96.25±0.64	95.27±1.11	95.75±0.88
PCA(5)+Deep-Forest	60.71±3.59	60.71±3.59	56.13±4.20	57.28±4.07
LDA(4)+Deep-Forest	91.25±1.27	91.25±1.27	87.20±2.71	89.83±1.63
PCA(5)+LDA(4)+Deep-Forest	79.64±0.73	79.64±0.73	75.33±1.70	77.08±1.04
Manual(5)+Deep-Forest	75.89±0.79	75.89±0.79	65.93±1.24	67.97±1.39

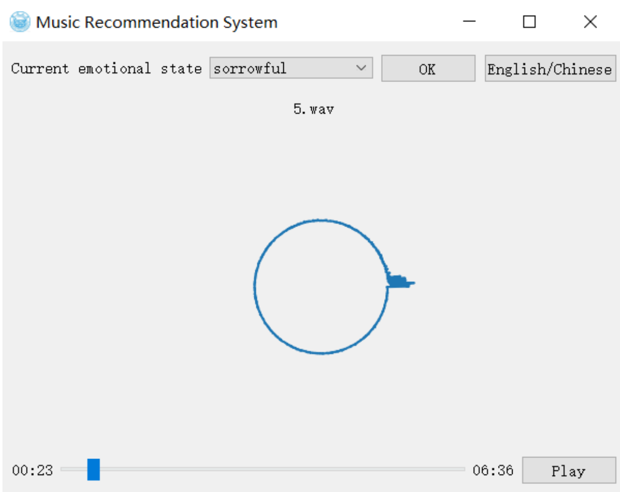

Figure 2. The protype of the user interface TJ-ERMR.

4. Conclusions

In this work, five different categories of emotion regulation music were selected which have been proved to be useful in music therapy. Using the Librosa library, 54 audio features of those music were extracted. Five significant audio features of emotion regulation music were found by manual selection and proved to have good classification performance across the five music categories. Finally, a protype of the emotion regulation music recommendation (TJ-ERMR) was built. In the future, more tracks can be added to this repertoire of emotion regulation music recommendation automatically using our classification model. This work opens up the prospect for more widespread adoption of music therapy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 92046008), the Shanghai Innovation Action Project of Science and Technology (No. 20511100700), and the Fundamental Research Funds for the Central Universities.

References

[[1] C. Yu-kun *et al.*, "Study of five elements music tracks and their clinical application in the treatment of depression " *China Journal of Traditional Chinese Medicine and Pharmacy,* vol. 34, no. 09, pp. 4234-4237, 2019.

[2] X. Ji-zong and L. Yue-ming, "The Principle for Sequence of Repertoire of Music Therapy with Experimental Observation," *HUANGZHONG(Journal of Wuhan Conservatory of Music, China),* no. 01, pp. 108-111, 2012.

[3] S. Uhlig, A. Jaschke, and E. Scherder, "Effects of music on emotion regulation: A systematic literature review," in *The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15*, 2013: University of Jyväskylä, Department of Music.

[4] S. Y. Jazi, M. Kaedi, A. Fatemi, and Applications, "An emotion-aware music recommender system: bridging the user's interaction and music recommendation," *Multimedia Tools,* vol. 80s, no. 9, pp. 13559-13574, 2021.

[5] K.-t. Song and C. Cervantes, "Music playing system and music playing method based on speech emotion recognition," ed: Google Patents, 2017.

[6] J.-z. XU and Y.-m. LI, "The Principle for Sequence of Repertoire of Music Therapy with Experimental Observation [J]," *HUANGZHONG(Journal of Wuhan Conservatory of Music, China),* vol. 1, 2012.

[7] B. McFee *et al.*, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, 2015, vol. 8, pp. 18-25: Citeseer.

[8] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *The Journal of the Acoustical Society of America,* vol. 63, no. 5, pp. 1493-1500, 1978.

[9] C.-h. Chen, *Signal processing handbook*. CRC Press, 1988.

[10] R. N. Shepard, "Circularity in judgments of relative pitch," *The journal of the acoustical society of America,* vol. 36, no. 12, pp. 2346-2353, 1964.

[11] M. L. Waskom, "Seaborn: statistical data visualization," *Journal of Open Source Software,* vol. 6, no. 60, p. 3021, 2021.

[12] P. Cunningham, "Dimension reduction," in *Machine learning techniques for multimedia*: Springer, 2008, pp. 91-112.

[13] M. Ringnér, "What is principal component analysis?," *Nature biotechnology,* vol. 26, no. 3, pp. 303-304, 2008.

[14] A. J. Izenman, "Linear discriminant analysis," in *Modern multivariate statistical techniques*: Springer, 2013, pp. 237-280.

[15] D. Hamilton *et al.*, "kNN vs. SVM: A Comparison of Algorithms," *Hood, Sharon M.,* vol. 78, pp. 95-109, 2020.

[16] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks,* vol. 10, no. 5, pp. 988-999, 1999.

[17] R. Thiruvengatanadhan, P. Dhanalakshmi, and P. S. Kumar, "Speech/music classification using SVM," *International Journal of Computer Applications,* vol. 65, no. 6, 2013.

[18] P. Dhanalakshmi, S. Palanivel, and V. Ramalingam, "Classification of audio signals using SVM and RBFNN," *Expert systems with applications,* vol. 36, no. 3, pp. 6069-6075, 2009.

[19] Z.-H. Zhou and J. Feng, "Deep forest," *arXiv preprint arXiv:.08835,* 2017.

[20] H. Wang, Y. Tang, Z. Jia, and F. Ye, "Dense adaptive cascade forest: a self-adaptive deep ensemble for classification problems," *Soft Computing,* vol. 24, no. 4, pp. 2955-2968, 2020.

[21] J. Cheng *et al.*, "Emotion recognition from multi-channel eeg via deep forest," *IEEE Journal of Biomedical Health Informatics,* 2020.

[22] M. W. Browne, "Cross-validation methods," *Journal of mathematical psychology,* vol. 44, no. 1, pp. 108-132, 2000.

[23] N. Ghamrawi and A. McCallum, "Collective multi-label classification," in *Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 195-200, 2005.