International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 12 (2018) pp. 10672-10676

Influence of Guided Particle Swarm Optimization in Automatic Music Emotion Recognition: A Comparative Study Using Different ANN Architectures

Nurlaila Rosli*, Nordiana Rajaee*

**Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.*
**Corresponding Author*

Abstract

The fundamental operations in soft computing and data mining, has given massive contribution in the development of electronic applications and intelligent system which in recent years, deliberately emphasized on the verbal information such as speech and music. This study presents the development of the music emotion recognition (MER) system that is able to classify 4 different emotions using the low level musical features taken from the Malay music dataset. The artificial neural network (ANN) is used throughout this study as a machine classifier incorporated with guided particle swarm optimization (GPSO) for the purpose of data training. The comparative test using different ANN architectures, with and without the GPSO is proposed as to investigate the impact of combining both algorithms towards the system performance. The results simulate that by incorporating GPSO with the ANN, the classification accuracy can be enhanced up to 90% and more. It is also proved that using GPSO instead of using the conventional PSO techniques somehow improved the musical features learning phase and leads to optimal MER performance.

Keywords: particle swarm optimization, music emotion recognition, artificial neural network

INTRODUCTION

Music has reached almost every single individual and can be access anytime, anywhere at any place. The escalating capacities of electronic data storage and affordable yet accessible digital music data via online streaming, has contributes to the growth of up to a few millions of digitally recorded data every single day [1]. With such massive numbers of music produce, it is not easy to search the music that suits our personal choice. Usually, music is organized and categorized based on specific class namely, artist, year, album, song title, genre, and also emotion.

Emotion based music classification, is still in the early stages, though in recent years, it has gained a lots of attention for a wide range of research area that is related to the subjects [2]. Until now, there are already various algorithms, machine learning, as well as data mining techniques, which has been exploits as to classify and recognize the association of music with moods or emotions. However, the predicament always transpire on how to accurately associates music with emotion? What are the criteria that must be considered in MER system that allows the system to intelligently categorize selected music data into certain emotion classes?

In the past 10 years, researcher in soft computing and data mining techniques have discovered wide improvement in analyzing and learning, not only the simple traditional data but also the musical audio data. Until now, the improvisation of training, recognizing, and machine learning techniques used in numbers of automatic music emotion recognition (MER) system has somehow increased the system performance as well as the classification accuracy.

Generally, there are three main important processes that must be considered during MER system development. First, identifying and defining criteria for the emotional model. Second, identifying and extracting audio features that associate music with emotion using signal processing algorithm. Third, train the audio data by selecting the most suitable and accurate classification algorithms and machine classifier.

Most of MER has automatically carried out using artificial intelligence (AI) machine classifier such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Particle swarm optimization (PSO), Decision Tree, fuzzy logic and etc [3]. In work [4], extracted music features are recognized using ANN and SVM model. Each model performed training and testing separately. The percentages of recognizing five main emotions in selected music are differ from each model. ANN outperforms SVM model up to 94.4% while SVM model achieved 85.0% of recognition performance.

Based on work in [5], simple backpropagation algorithm with multilayer's ANN trained many times to produce 94.4% classification accuracy for 8 types of emotions in Indian popular music. According to Chiang et al. in [6], the average classification accuracies using hierarchical SVM's classifier for two music datasets are 86.94% and 92.33% respectively. Overall, the effectiveness of the proposed techniques and algorithm in most automatic MER was validated by the positive outcomes. Thus it is crucial to select the most appropriate machine classifier for a better MER performance.

The rest of this paper is organized as follows; in section 2, we illustrate the proposed machine classifier used in this research, including the introduction and justification of ANN architecture and GPSO learning techniques. Section 3 presents the whole comparative experiment preparation. The results and relative discussion will be presented in section 4. We conclude with an overview of MER systems and suggestion for future work.

MUSIC EMOTION RECOGNITION CLASSIFIER

In this section, we describe the machine classifier proposed and used throughout this research. According to work in [7], there are four important factors that impact the classification success namely; features format, features selection, training data, and machine classifier. However, this paper highlighted only on the machine classifier factors. The purpose of this research is to mine the impact of using different ANN architecture toward classification accuracy. We also incorporated GPSO with ANN as to compare the end results to choose the best model classifier for a better MER system performance.

Artificial Neural Network (ANN) Architecture

Three types of ANN architecture will be exploits in this research along with guided particle swarm optimization. Details justifications are explained in the subsection below.

Feed Forward Backpropagation Neural Network (FFBPNN)

Feed-forward backpropagation neural networks (FFBPNN) involves both feed-forward and backpropagation algorithm. During training session, the system will calculate the error, which is defined as the square of the difference between the actual and the desired activities [8] [9].In FFBPNN, back propagation is the algorithm that computes the error weight derivative (EW) as to reduce error during training and testing. The EW calculated proportionally to the rate at which the error and the weight changes. The changes of EW will also transform the activity level of a unit (EA) which imply the value of difference between actual and desire output.

To compute EA, all connected weight for hidden and output units must be identified first. Those weights multiply by EAs of those output units and add the products. For the sum equals the EA for the chosen hidden unit. All step must be repeated from layer to layer in which the activities propagate through the network and from opposite direction. Finally, the EW for each connection can be identified as EW is the product of EA and activity through the incoming connection.

Radial Basis Function Neural Network (RBFNN)

The idea of Radial Basis Function Neural Network (RBFNN) derives from the theory of function approximation. RBF Networks take a slightly different approach with much appropriate features compare to other network. Basically RBFNN architecture is build up with two layer feed-forward, hidden nodes with radial basis function and output nodes implement linear summation function same as multilayer perceptron (MLP). According to works[10]–[12] justification of RBFNN layers are shown in Table 1.

Table 1. Justification of RBFNN Architecture

RBFNN Architecture	Justification
Input Layer	Consist of one neuron for each predictor variable. For categorical variable, N represents the categories and N-1 is used as a processing value before the input layer. The input neuron computed by subtracting the median and dividing by the interquartile range and then feed to each neuron in the hidden layer.
Hidden Layer	For hidden layer the optimal number is determined by the training process. The predictor variable for this hidden unit comes with radial basis function (RBF). Each RBF may be different based on the dimensions. The centres and spreads depend on the training process. When X vector represent the input value for input layer, Euclidean distance will be calculated in hidden layer, the RBF kernel then will be applied to this distance by using the spread value. The result then passed to the summation layer.
Output Layer	To present as an output of the network, the value from the hidden nodes is multiplied by weight associated with neuron (W1, W2..Wn) and assed to summation. Bias value on the other hand is the value between 1.0 multiplied by a weight W0 and feed into submission layer. For classification problems, there is one output and a separate set of weights and summation unit for each target category. The value output for a category is the probability that the case being evaluated has that category.

Fully Recurrent Neural Network (FRNN)

FRNN is the basic architecture developed in the 1980s. Consist of input, hidden, and output nodes. FRNN neuron work in unit with each neuron connected directly to every other neuron. Each neuron has a time-varying real-valued (more than just zero or one) activation (output) [9]. Weight for each connection altered to activate the input nodes by setting up the real input value vector during the training. Each non input unit computes its current activation which receives connections. There may be teacher-given target activations for some of the output units at certain time steps.[13][14].

Guided Particle Swarm Optimization (GPSO)

Basically particle swarm optimization PSO is inspired by the nature social behavior and dynamic movement with communications of, insects, birds and fish [15]. Three simple behaviors namely, separation, alignment and cohesion are implying in the PSO algorithm based on self and social experiences.

Collection of changing particles leads to changing solution. Possible solution is depends on the search area by moving towards promising area and get global optimum. Each particle must keep track of its personal best, *pbest* value and global best, *gbest* value. GPSO in other hand is the improvement algorithm taken from the PSO improvisation based on problem solution and size of particle. In this research GPSO is used to train audio features that have been extracted from Malay music data. The movement of PSO is guided using intensification and diversification formula in (1).

$$\mathbf{v}_i^{t+1} = \underbrace{\mathbf{v}_i^t}_{\text{Diversification}} + \underbrace{\mathbf{c}_1\mathbf{U}_1^t(\mathbf{pb}_i^t - \mathbf{p}_i^t) + \mathbf{c}_2\mathbf{U}_2^t(\mathbf{gb}^t - \mathbf{p}_i^t)}_{\text{Intensification}} \quad (1)$$

Intensification allows particles to identify the solution from the previous stage, find the best solution of a given region, meanwhile *diversification*, is a process on searching new solutions, finds the regions with potentially the best solution[16]. The GPSO algorithm consists of:

- Particle Evaluation
- Set the individual and global best fitness and position
- Modify particles velocity and position.

COMPARATIVE EXPERIMENT PREPARATION

Generally, the algorithms proposed in most MER system are typically similar to one another. In the initiation stage audio data needs to be standardizing into the same format such as Wav and split into 30 sec frame data. According to numbers of researcher in emotion classification in music including [17], [18], [19], [10] the reasons of using 15-30 sec audio data is to avoid wrong features extraction which may leads to inaccuracy of training and testing result.

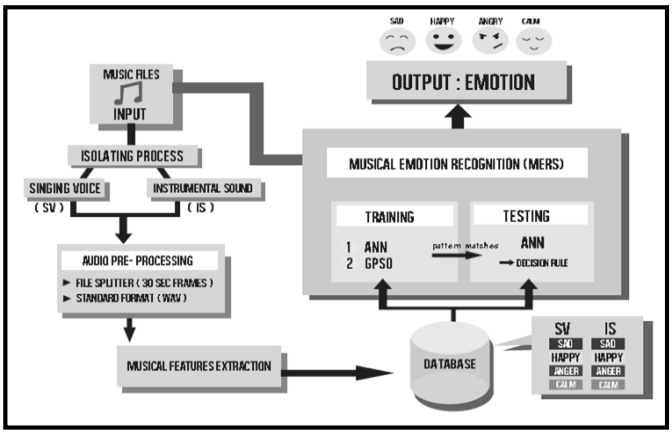

Figure 1. Overall processes in music emotion recognition (MER) system.

Features Selection and Extraction

Overall seven musical features are selected based on its attributes that are strongly correlated with specific emotion. MIRToolbox and MA Toolbox in Matlab is used to extract all seven features namely; *zerocross, rms, spectral centroid, spectral rooloff, mfcc, tempo and irregularity*. It has been proved that all seven features used in this research are very practical in exemplifying divergent audio signals and have been universally used in speech and music classification problems [20], [21].

Test Setup

This test involves six different algorithms. This is to compare the differences in classification rate and system performance when different algorithm is exploits. Matlab programming language is used throughout this process to build training and testing model for each algorithm. The details of algorithm used are as follow:

- FFBPNN: Audio features data extracted are trained and tested using feed-forward backpropagation neural network.
- RBFNN: Audio features data extracted are trained and tested using gaussian function in radial basis function neural network.
- FRNN: Audio features data extracted are trained and tested using fully recurrent neural network
- FFBPNN + GPSO: Extracted audio data features are trained using GPSO and tested using feed-forward backpropagation neural network.
- RBFNN +GPSO: Extracted audio data features are trained using GPSO and tested using radial basis function neural network
- FRNN +GPSO: Extracted audio data features are trained using GPSO and tested using fully recurrent neural network

EXPERIMENTAL RESULTS AND DISCUSSION

The experimental comparison using different ANN architectures and different algorithm by incorporating GPSO for data learning and training has been presented in this paper.

Classification Results

The result accuracy are measured using the standard performance measurement based on work done in[22]. The higher the percentage, the better the performance and classification accuracy. 200 music data that has been manually annotated accordingly into 4 main emotions are used for system testing. Results for classification accuracy using different algorithm are shown in Table 2.

$$\text{Accuracy \%} = \frac{\text{Numbers of Correctly Classified Music Data}}{\text{Total Numbers of Music Data}} \times 100$$

Figure 2. Standard measurement to measure the percentages of classification accuracy

Table 2. Classification accuracy using different algorithm

Algorithm	ANN Architectures	Classification Accuracy %
With GPSO	FFBPNN	81.7
	RBFNN	93.3
	FRNN	73.1
Without GPSO	FFBPNN	71.6
	RBFNN	84.9
	FRNN	65.3

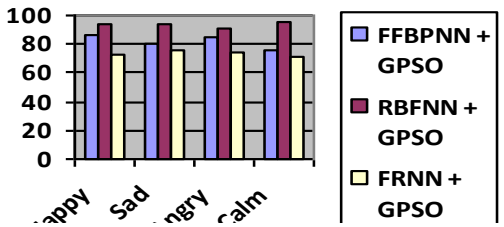

Figure 3. Percentages of classification accuracy by combining ANN with GPSO

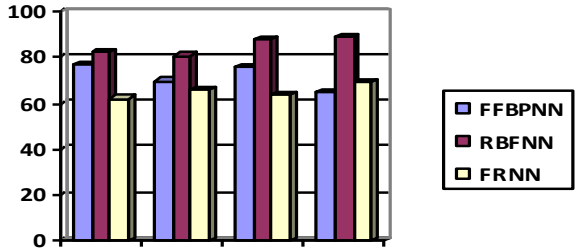

Figure 4. Percentages of classification accuracy using ANN model classifier without GP

RESULTS DISCUSSION

The algorithm is created to classify emotion in selected music by comparing happy, anger, calm and sad audio data features that have been successfully extracted and grouped based on type of emotion. The network was built, trained and tested using the MATLAB programming language. All of musical data are successfully processed and extracted from 1000 Malay music database. Both ANN and GPSO training model have systematically trained all 800 audio before the ANN testing model is then used to test the selected 200 music from all four emotions. At the end, MER systems with the combination of ANN and GPSO learning algorithm have successfully classified emotion in selected songs.

CONCLUSION AND FUTURE WORKS

In this work, we present a comparison study evaluation of total six different algorithms, for automatic music emotion recognition. Seven low-levels audio features in each input music files are extracted and trained using ANN as well as combination of both ANN and GPSO classifiers as to compare the overall classification performance. Around 1000 Malay popular songs that has been manually annotated and grouped accordingly into four main emotions have been used as a training data throughout this study. About 800 music data are used for training purposes and the other 200 music data are used for testing.

ANN and GPSO model classifier is build using Matlab programming language. ANN model training, trained all the musical features extracted from 800 music data. All trained parameters have been loaded into the database for testing. The same operations repeat using the combination of both methods (ANN+GPSO). The final system is able to associates music with specific emotions by comparing the features parameters in the database and thus classified music based on specific emotion.

Generally, the algorithm developed is proven to be up to 90% accurate. Results shows that the combination of RBFNN with GPSO achieved highest classification accuracy to be exact 93.3% compare to other algorithm. However, by using FRNN architecture, the system only able to classify 32 music or more than half of the total music data which make it capitulate the lowest classification rate 65.3%. It is also proved that by incorporating ANN and GPSO, the percentages of classification accuracy increase accordingly for each algorithm.

For future improvement, it is suggested that others machine learning techniques such as SVM, fuzzy logic, decision three, genetic algorithm are exploits along with GPSO to deliberately find the most fitting combination for a better MER system performance.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Malaysia Sarawak for supporting the research. Grant Number: F02(DPI28)/1244/2015(02).

REFERENCES

[1] Alexandridis A., Chondrodima E., Paivana G., Stogiannos M., Zois E., and Sarimveis H., (2014), Music Genre Classification Using Radial Basis Function Networks and Particle Swarm Optimization, Computer Science and Electronic Engineering Conference (CEEC), 35 – 40.

[2] Kim Y.E., Schmidt E.M., Migneco R., Morton B.G., Richardson P., Scott J., Speck J.A., and Turnbull D., (2010). Music Emotion Recognition : A State Of The Art Review, 11th International Society for Music Information Retrieval Conference (ISMIR), 255–266.

[3] Nalini S., Palanivel N.J., (2013), Emotion Recognition In Music Signal Using Ann And Svm, Int. J. Comput. Appl., Vol. 77, No. 2, 7–14.

[4] Chen, L., Mao, X., Xue, Y., & Cheng, L. L. (2012). Speech emotion recognition: Features and classification models. Digital signal processing, 22(6), 1154-1160.

[5] Bhat A.S, Amith V.S., Prasad N.S., and Mohan D.M., (2014), An Efficient Classification Algorithm For Music Mood Detection In Western And Hindi Music Using Audio Feature Extraction, 2014 Fifth Int. Conf. Signal Image Processing, 359–364.

[6] Chiang W.C, Wang J.S, and Hsu Y.L, (2014), A Music Emotion Recognition Algorithm With Hierarchical Svm Based Classifiers, 2014 Int. Symp. Comput. Consum. Control, 1249–1252.

[7] Zentner, M., Grandjean, D., and Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion, 8(4), 494.

[8] Benardos, P. G., and Vosniakos, G. C. (2007). Optimizing feedforward artificial neural network architecture. Engineering Applications of Artificial Intelligence, 20(3), 365-382.

[9] Zhang, C., Shao, H., and Li, Y. (2000). Particle swarm optimisation for evolving artificial neural network, IEEE International Conference on Systems, Man, and Cybernetics, 2487-2490

[10] Erol, R., Ogulata, S. N., Şahin, C., and Alparslan, Z. N., (2008). A radial basis function neural network (RBFNN) approach for structural classification of thyroid diseases. Journal of medical systems, 32(3), 215-220.

[11] Yingwei, L., Sundararajan, N., and Saratchandran, P. (1998). Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm. IEEE Transactions on neural networks, 9(2), 308-318.

[12] Korürek, M., and Doğan, B. (2010). ECG beat classification using particle swarm optimization and radial basis function neural network. Expert systems with Applications, 37(12), 7563-7569.

[13] Williams, R. J., and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. Neural computation, 1(2), 270-280.

[14] Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. IEEE transactions on neural networks, 5(2), 240-254.

[15] Shi, Y., and Eberhart, R. C. (1999). Empirical study of particle swarm optimization.IEEE Proceedings of the 1999 congress on In Evolutionary computation, (3), 1945-1950.

[16] Riget, J., and Vesterstrøm, J. S. (2002). A diversity-guided particle swarm optimizer-the ARPSO. Dept. Comput. Sci., Univ. of Aarhus, Aarhus, Denmark, Tech. Rep, 2.

[17] Eerola T., Lartillot O., and Toiviainen P. (2009). Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models, ISMIR, 621-626.

[18] Cabredo R., Legaspi R. S., Inventado P. S., and Numao, M. (2012). An Emotion Model for Music Using Brain Waves, ISMIR, 265-270.

[19] Hu, X., Downie, J. S., and Ehmann, A. F. (2009). Lyric text mining in music mood classification. American music, 183(5,049), 2-209.

[20] Yang Y. H., Lin Y. C., Cheng H. T., Liao I. B., Ho Y. C., and Chen, H. H. (2008). Toward multi-modal music emotion classification. In Pacific-Rim Conference on Multimedia, 70-79.

[21] Li T., and Ogihara M. (2004). Content-based music similarity search and emotion detection, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings, 5,705-8d