Music Genre and Emotion Recognition Using Gaussian Processes and Neural Network

Femy Sunny[1], Sreevarsha V[2], Jamseera K[3], Nijisha P[4]

[1,2,3,4]*Department of electronics and communicationengineering,Cochin college of engineering,valanchery*

ABSTRACT

The main objective of this work is to develop a music emotion recognition and genre classification method by using Gaussian process and neural network. The traditional methods for the above tasks are make use of support vector machine and two dimensional valence arousal models. To achieve high accuracy, here we introducing a new method using Gaussian process with neural network. Gaussian process is flexible to any type of kernel and its deals uncertainty very efficiently. The dataset required are also less. The music information is determined by extracting the features such as Mel-frequency cepstral coefficients (MFCC), timbre features, spectral crest factor, spectral flatness measure. Then we applied the neural network on the above features to find out the genre and emotion of the music. The experimental results showed that the proposed Gaussian process with neural network system performed consistently better than the neural network in both music emotion recognition and genre classification tasks.

Keywords: music emotion recognition, music genre classification, Gaussian process, neural network, Mel-frequency cepstral coefficients, timbre features, spectral crest factor, spectral flatness measure

I. INTRODUCTION

Most of people enjoy music in their free time and it will help them to reduce their metal stress. Almost all music is created to convey emotion to the listeners. Now a day, a lots of music files are available locally or over the internet. To structure and organize the large number of music files available digitally on the web, we need an efficient music information retrieval system. The music genre classification and the emotion recognition are two major tasks in music information retrieval.

Music genres were labels created and used by humans for describing and categorizing the world of music. It has no strict definitions. While in the case of music emotion recognition there are many issues for the reasons that, the same song can generate different emotions in different persons according to different situations and it is still very strange how music evokes emotion. The manual genre classification and emotion recognition can be inconsistent and changing. Because the human judgements are influenced by artist fashion, dance styles, lyrics, social and political attachments, religious believes, etc. So we require automatic system for music genre classification and the emotion recognition.

Here we are investigating the applicability of Gaussian process model for music genre classification and emotion recognition tasks and comparing their performance with neural network. The ultimate goal of this paper is to propose an efficient system for above tasks with high accuracy.

The paper is structured as follows. Section II describes the proposed system, section III Gaussian process, Section IV feature extraction and Section V classification. Finally section VI with conclusions.

II.SYSTEM DESCRIPTION

In the proposed method for music genre classification and emotion recognition using Gaussian process and neural network, first of all, we pre-emphasis the music signal. This process is to increase the magnitude of higher frequencies with respect to the magnitude of the lower frequencies. It will help to increase the energy of the signal at higher frequency, as they weak in music signal. Extract the features of the input signal. The extracted features are given to normalization and ranking. Normalization is a process to avoid duplicate and redundant data and in ranking matrix of feature is converted into the same rank. The feature extraction and classification modules are same for both music genre classification and emotion recognition. For classification task here we use Gaussian process and neural network.

III. GAUSSIAN PROCESS

In probability theory and statistics, a family of statistical distributions (not necessarily stochastic processes) in which time have an important role, is known as Gaussian processes. Gaussian processes were used to describe distributions over functions. The GP is defined as a collection of random variables of any finite number which has a joint Gaussian distribution. Gaussian processes are completely specified by its mean and covariance functions. For a real process f(x), the mean function m(x) and the covariance function k(x, x') are defined as,

$$m(x) = \mathbb{E}[f(x)]$$
$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$$

Thus, the GP can be represented as

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

A GP prior over function $f(x)$ implies that for any finite number of inputs X ={xi } ∈ $\mathbb{R}d$, i =1,…,n, the vector of function values $f = [f(x1), \dots, f(xn)]^T = [f1, \dots, fn]^T$ has a multivariate Gaussian distribution

$$f \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$$

Where, the mean $\boldsymbol{m}$ is often assumed to be zero. The form of covariance matrix $\boldsymbol{K}$ has given,

$$K = \begin{bmatrix} K(x_1,x_1) & \dots & K(x_1,x_n) \\ K(x_2,x_1) & \dots & K(x_2,x_n) \\ & \vdots & \\ K(x_n,x_1) & \dots & K(x_n,x_n) \end{bmatrix}$$

and characterizes the correlation between different points in the process. For (x, x'), any kernel function which produces symmetric and semi-definite covariance matrix can be used.

IV. FEATURE EXTRACTION

The process of computing a compact numerical representation that can be used to characterize a segment of audio is known as feature extraction. Feature extraction from the audio signal is very important in the sense that they represent the music well and computation can be carried out efficiently. The different features extracted in this work are mel frequency cepstral coefficients, zero crossing rate, spectral centroid, spectral flux, spectral rolloff, spectral crest factor, chromagram, linear prediction spectral coefficients. And the extracted features are stacked into a feature vector. This feature vector is used for the further calculations.

V. CLASSIFICATION

Artificial neural networks have become innovative classification methods which imitate the learning process same as the human brain. In this sense, the information processing structure of an artificial neural network is composed of a large number of highly interconnected processing elements working in parallel to solve specific problems. The network is constructed by an input layer, one or more hidden or middle layers, and an output layer. The number of units in the hidden layers are variable, the number of units in the input and output layer are fixed by the problem conditions. The input layer it is defined by the number of variables or features used to represent the data samples, and in the output layer it is defined by the number of output or classes required.The network was built, trained and tested by using the MATLAB programming language. Neural network toolbox in MATLAB was utilized to train the neural network.

VI. CONCLUSION

The systems were implemented in MATLAB to get the result for music genre classification and emotion recognition. Gaussian process will provide probabilistic prediction and gives an estimate of uncertainty in the prediction. Neural nets offer massive parallelism for real-time operation and adaptation, which has the

potential of helping to solve difficult classification tasks. Here we consider music files from ten different genres and six different emotions. The system developed is proven to be up to 98.9% accurate for music genre classification and 96.8% for music emotion recognition. The performance of music genre classification and emotion recognition is improved by combining both Gaussian process and neural network.

REFERENCES

[1]. George Tzanetakis, Perry Cook "Musical Genre Classification of Audio Signals" ieee transactions on speech and audio processing, vol. 10, no. 5, July 2002.

[2]. Yi-Hsuan Yang, Homer H. Chen "Prediction of the distribution of perceived music emotions using discrete samples" ieee transactions on audio, speech, and language processing, vol. 19, no. 7, September 2011

[3]. Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang "Survey Of Audio-Based Music Classification And Annotation" ieee transactions on multimedia, vol. 13, no. 2, April 2011

[4]. Lie Lu, Dan Liu, and Hong-Jiang Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals" ieee transactions on audio, speech, and language processing, vol. 14, no. 1, January 2006

[5]. Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su,Homer H. Chen, "A Regression Approach to Music Emotion Recognition" ieee transactions on audio, speech, and language processing, vol. 16, no. 2, February 2008

[6]. K. Markov, M. Iwata, and T. Matsui, ``Music emotion recognition using Gaussian processes," in Proc. 2nd ACM Multimedia Workshop Crowd-sourcing Multimedia, Barcelona, Spain, Oct. 2013.

[7]. M. Soleymani, M. N. Caro, E. M. Schmidt, C. Y. Sha,Y. H. Yang, ``1000 songs for emotional analysis of music," in Proc. 2nd ACM Multimedia Workshop Crowd sourcing Multimedia, Barcelona, Spain, 2013.

[8]. T. Li and M. Ogihara, ``Detecting emotion in music," in Proc. Int. Soc. Music Inform. Retr. Conf (ISMIR), vol. 3. Oct.2003, pp. 239_240.

[9]. M. Barthed, G. Fazekas, M. Sandler, ``Multidisciplinary perspectives on music emotion recognition: Implications for content and context based models," in Proc. 9th Symp. Comput. Music Model. Retr.(CMMR), Jun. 2012, pp. 492_507.

[10]. R. Typke, F. Wiering,R. C. Veltkamp, ``A survey of music information retrieval systems," in Proc. Int. Conf. Music Inform. Retr., 2005, pp. 153_160.

[11]. I. Panagakis, C. Kotropoulos, and G. Arce, ``Music genre classification using locality preserving non-negative tensor factorization and sparse representations," in Proc. Int. Soc. Music Inform. Retr. Conf.(ISMIR), 2009, pp. 249_254.

[12]. M. Ebden, "Gaussian Processes for Regression: A Quick Introduction" August 2008

[13]. Mudiana Binti Mokhsin, Nurlaila Binti Rosli,cSuzana Zambri, Nor Diana Ahmad, Saidatul Rahah Hamidi "Automatic music emotion classification using Artificial neural network based on vocal and Instrumental sound timbres" Journal of Computer Science 10 (12): 2584-2592, 2014

[14]. Saravanan K, S. Sasithra "Review on classification based on artificial Neural networks" International Journal of Ambient Systems and Applications (IJASA) Vol.2, No.4, December 2014

[15]. Kleanthis Koupidis "Emotion recognition and classification using neural networks" Master thesis, Mathematics Department, Aristotle University of Thessaloniki, June 2014

[16]. Naresh N. Vempala and Frank A. Russo "Predicting Emotion from Music Audio Features Using Neural Networks" 9th International Symposium on Computer Music Modelling and Retrieval (CMMR), Queen Mary University of London 19-22 June 2012.

[17]. Braja Gopal Patra, Promita Maitra, Dipankar Das and Sivaji Bandyopadhyay "Music Emotion Recognition based on Feed-Forward Neural Network" MediaEval 2015 Workshop, September 14-15, 2015, Wurzen, Germany.

[18]. N.J. Nalini, S. Palanivel "Emotion Recognition in Music Signal using AANN and SVM" International Journal of Computer Applications (0975 – 8887) Volume 77– No.2, September 2013

[19]. Gao Feng, Konstantin Markov, Jianguo Yu, "Music Emotion Recognition" Proceedings of the 2nd International Conference on Applications in Information Technology (ICAI)'16, Oct.6–8,2016, Aizu-Wakamatsu,Japan.