IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Audio based Music Classification based on Genre and Emotion using Gaussian Process

Mugdha Magare[1], Prof. Ranjana Dahake[2]

ME Student, Dept of Computer Engineering, MET BKC, University of Pune, Nashik, India[1]

Professor, Dept of Computer Engineering, MET BKC, University of Pune, Nashik, India[2]

Abstract: In the field of Music Information Retrieval (MIR), music genre classification and music emotion recognition are the two main tasks to investigate for further development. In this project work, these two tasks are focused. For this purpose, Gaussian Processes Model is used. Gaussian Processes (GPs) are Bayesian nonparametric models that are becoming more popular for their superior capabilities to capture highly nonlinear data relationships in various tasks, such as dimensionality reduction, time series analysis, novelty detection, as well as classical regression and classification tasks. Gaussian Processes are used to investigate the feasibility and applicability of Gaussian Process model for music genre classification and emotion estimation. Along with this, we are reducing the time required for feature extraction for classification tasks. Principle component analysis (PCA) technique is used to reduce this time. In this, it selects and considers only higher order features for classification.

Keywords: Music Information Retrieval, Genres, Emotions, Features.

I. INTRODUCTION

Now days, thousands of music files are uploaded and downloaded through the internet. Efficient retrieval of these audio files is necessity of such users who are willing to get access to these music files. Lot of research is going on for the efficient music information retrieval technology. Research in this area has focus on tasks such as genre classification, artist identification, mood estimation, cover song identification, music annotation, melody extraction, etc. All these tasks are useful for efficient music search and recommendation services, playlist generation and other applications. Information sources for MIR can be: text based - music related Internet sites, social networks, lyrics, etc; and audio based - the music signal itself or mixed text and audio. Here, audio based information source is going to be considered for music genre classification and music emotion estimation tasks.

Musical genres are labels created and used by humans for categorizing and describing the vast universe of music. Musical genres have no strict definitions and boundaries as they arise through a complex interaction between the public, marketing, historical, and cultural factors. This observation has led some researchers to suggest the definition of a new genre classification scheme purely for the purposes of music information retrieval. However even with current musical genres, it is clear that the members of a particular genre share certain characteristics typically related to the instrumentation, rhythmic structure, and pitch content of the music. Automatically extracting music information is gaining importance as a way to structure and organize the increasingly large numbers of music files available digitally on the Web. It is very likely that in the near future all recorded music in human history will be available on the Web. Automatic music analysis will be one of the services that music content distribution vendors will use to attract customers.

Genre classification is a classical supervised classification task. The goal is to predict the genre of an unlabelled music piece. Human categorized music by genres seems inconsistent and keeps changing by time. It depends on human judgements which are influenced by many factors such as audio signal artist fashion, dance style, lyrics, social and political attachment, etc. By the time new genres are arriving continuously. Thus it’s impossible to come up with commonly agreed set of music genres.

Although users are more likely to use genres or artists names when searching or categorizing music, the main power of music is in its ability to communicate and trigger emotions in listeners. Thus, determining computationally the emotional content of music is an important task. Music itself is the expression of emotions, which can be highly subjective and difficult to quantify. Automatic recognition of emotions (or mood) in music is still in its early stages, though it has received increasing attention in recent years. Determining the emotional content of music audio computationally is, by nature, a cross disciplinary endeavour spanning not only signal processing and machine learning, but also requiring an understanding of auditory perception, psychology, and music theory.

II. RELATED WORK

Automatically extracting music information is gaining importance because of a need to organize the increasingly large numbers of music files available digitally on the

DOI 10.17148/IJARCCE.2016.5795

Web. It is very likely that in the near future all recorded music in human history will be available on the Web.
Musical genres are labels created and used by humans for categorizing and differentiating music. Musical genres have no fix definitions. It also does not have boundaries as they arise through interaction between the public, marketing, historical, and cultural factors [2]. Genre hierarchies, typically created manually by human experts, are currently one of the ways used to structure music content on the Web. Automatic musical genre classification will automate this process and provide an important component for a complete music information retrieval system for audio signals. In addition it provides a framework for developing and evaluating features for describing musical content. Such features can be used for other music information retrieval tasks and form the foundation of most proposed audio analysis techniques for music.

Three feature sets for representing timbral texture, rhythmic content and pitch content of music signals were proposed and evaluated by G. Tzanetakis & P. Cook using statistical pattern recognition classifiers trained with large real-world audio collections [2]. The proposed features sets successfully testified thus can be used in other music information retrieval tasks. Another approach presented was Compressing Sampling based approach [3]. In that, they present CS-based classifier for music genre classification, with two sets of features, including short-term and long-term features of audio music. The proposed classifier generates a compact signature to achieve a significant reduction in the dimensionality of the audio music signals.

Another faster approach to extract features is investigated by M. Henaff et.al. [4] In this, they investigated a sparse coding method called Predictive Sparse Decomposition (PSD) that attempts to automatically learn useful features from audio data. Due to its faster nature, it is scalable to large scaled datasets.
Music is composed to be emotionally expressive, and emotional associations provide an especially natural domain for indexing and recommendation in today's vast digital music libraries [5]. But such libraries require powerful automated tools, and the development of systems for automatic prediction of musical emotion presents a myriad challenges. The perceptual nature of musical emotion necessitates the collection of data from human subjects. The interpretation of emotion varies between listeners thus each clip needs to be annotated by a distribution of subjects.

E. Kim et al., made comparison in between state-of-the - art techniques for emotion recognition [5]. They have compared human annotations, contextual text information and content-based audio analysis. Still problems remain due to inherent ambiguities of human emotions. The VA plane which is generally used to point emotions on, it is proposed by James Russell [6]. It contains emotions placed on VA plane by their emotional characteristics.

M. Casey et.al. designed a CS-based classifier for music genre classification, with two sets of features, including short-time and long-time features of audio music. The proposed classifier generates a compact signature to achieve a significant reduction in the dimensionality of the audio music signals. They also outline the problems of content-based music information retrieval and explore the state-of-the-art methods using audio cues (e.g., query by humming, audio fingerprinting, content-based music retrieval) and other cues (e.g., music notation and symbolic representation), and identifies some of the major challenges for the coming years [7]. Studies in music processing have investigated various feature types and their extraction algorithms [9].

After the work of Rasmussen and Williams [11] which introduced GPs for the machine learning tasks of classification and regression, many researchers have utilized GPs in various practical applications. As SVMs, they are also based on kernel functions and Gram matrices, and can be used as their plug-in replacement. The advantage of GPs with respect to SVMs is that their predictions are truly probabilistic and that they provide a measure of the output uncertainty. Another big plus is the availability of algorithms for their hyper parameter learning.

However, recent studies have suggested that regression approaches using continuous mood representation can perform better than categorical classifiers [12]. In one of the earliest studies, features representing timbre, rhythm, and pitch have been used in SVM based system to classify music into 13 mood categories [13].

Markov and Matsui have proposed a novel approach for genre classification and emotion estimation [1], [10]. They have investigated the feasibility and applicability of Gaussian Process models for genre classification and emotion estimation on MediaEval'13. Gaussian Processes (GPs) are Bayesian nonparametric models. We are extending this work by creating own dataset and applying Gaussian process upon it. Also we are reducing time for feature extraction by PCA technique mentioned in [8]. Babu Kaji Baniya, Joonwhoan Lee and Ze-Nian Li have done analysis of audio features and reduction of them for automatic music genre classification using Principle Component Analysis (PCA) [8].

III.PROPOSED SYSTEM

In this system feature extraction is going to be performed. Both tasks i.e. Genre classification and Emotion recognition requires firstly feature extraction. Thus, it is obvious to have audio signal, the one and only, as an input to the system. Along with this we are reducing the time required to extract features by feature analysis and reduction technique. Dataset is used for classification also we are taking real world clips and classifying them. Real world clips are classified on the basis of training and testing mechanism.

DOI 10.17148/IJARCCE.2016.5795

A. SYSTEM FLOW

System flow shown in following figure 1 is explained in detail in following three topics respectively.

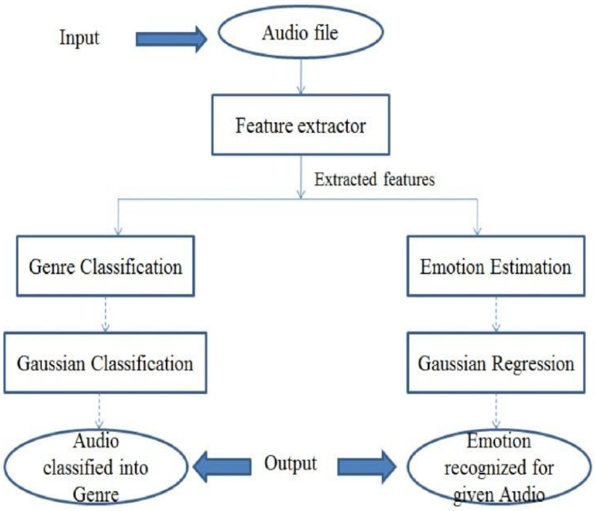

Figure 1: System Flow

A] Feature Extraction

In this work, used feature extraction methods in signal processing work are widely used and can be said as a standard set for these tasks. Features to be extracted are as follows:

- MFCC (Mel Frequency Cepstral Coefficients) - This is a feature widely used in automatic speech and speaker recognition.
- TMBR (timbre feature) - A set of four scalar features consisting of spectral centroid, spectral flux, spectral rolloff, and zero crossings.
- SCF and SFM (spectral crest factor and spectral flatness measure) - These features are subband based measures indicating spectral shape and used to discriminate between tone-line and noise-like sounds.
- CHR (chromagram) - This feature represents the spectrum distribution of the distinct semitones and provides information about the key & mode.

b] Genre Classification

Musical genres are categorical labels created by humans to characterize pieces of music. A musical genre is distinguished by the common characteristics shared by its members. These characteristics typically are related to the rhythmic structure, instrumentation, and harmonic content of the music. Genre hierarchies are commonly used to structure the large collections of music available on the Web. Currently musical genre annotation is carried out manually. Purpose is to reduce this manual work and make it automatic.

c] Emotion Estimation

The nature of human emotion is perceptual; it is a tough task to derive an intuitive and coherent set of adjectives and their specific grouping.

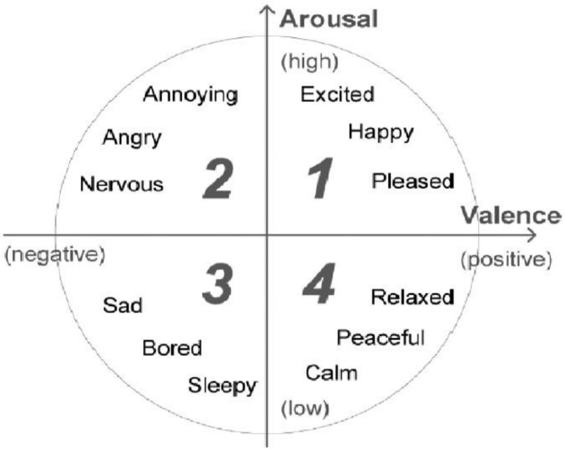

Figure 2: The Russell's two-dimensional Valence-Arousal (VA) space

To alleviate the challenge of ensuring consistent interpretation of mood categories, some studies propose to describe emotion using continuous multidimensional metrics defined on low-dimensional spaces. Most widely accepted and well known is the Russell's two-dimensional Valence-Arousal (VA) space where emotions are represented by points in the VA plane as shown in figure 2. It consists of the space where some regions are associated with distinct mood categories. Here, the task of music emotion recognition is to automatically find the point in the VA plane which corresponds to the emotion induced by a given music piece. Since the Valence and Arousal are by definition continuous and independent parameters, we can estimate them separately using the same music feature representation.

B. METHODOLOGY

Gaussian Processes (GP) are used to describe distributions over functions. The GP is defined as a collection of random variables any finite number of which has a joint Gaussian distribution. It is completely specified by its mean and covariance function. For real process f (x), the mean function m(x) and the covariance function k(x, x') are defined as,

$$m(x) = \mathrm{E}[f(x)] \qquad \text{eq. (1)}$$

$$k(x, x') = \mathrm{E}[(f(x) - m(x))(f(x') - m(x'))] \qquad \text{eq. (2)}$$

Thus, the GP can be written as,

$$f(x) \sim GP(m(x), k(x, x')) \qquad \text{eq. (3)}$$

A GP prior over function f (x) implies that for any finite number of inputs X = {x_i} ϵ R^d, i = 1,...,n, the vector of function values f = [f (x_1), ..., f (x_n)]^T = [f_1, ..., f_n]^T has a multivariate Gaussian distribution using eq. (1) and (2).

$$f \sim N(m, K) \qquad \text{eq. (4)}$$

DOI 10.17148/IJARCCE.2016.5795

where,
The mean **m** is often assumed to be zero. N is the multivariate Gaussian distribution. The covariance matrix **K** has the following form,

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

For Emotion estimation purpose, Gaussian process regression is used. And for genre classification purpose, gaussian process classification is used.

A] Gaussian Process Classification

For binary classification, given training data vectors $x_i \in R^d$ with corresponding labels $y_i \in \{-1, +1\}$, here to predict the class membership probability of a test point x*. This is done using an unconstrained latent function f (x) with GP prior and mapping its value into the unit interval [0, 1] by means of a sigmoid shaped function. It is carried out using logistic function.
Let $X = [x_1, \dots , x_n]$ be the training data matrix, $y = [y_1, \dots, y_n]^T$ be the vector of target values, and $f = [f_1, \dots, f_n]^T$ with $f_i = f(x_i)$ be the vector of latent function values. Given the latent function, the class labels are assumed independent Bernoulli variables and therefore the likelihood can be factorized as shown in eq. (5),

$$p(y \mid f) = \prod_{i=1}^{n} p(y_i \mid f_i) = \prod_{i=1}^{n} sig(y_i f_i) \qquad \text{eq. (5)}$$

b] Gaussian Process Regression

Given input data vectors $X = \{x_i\}$, i = 1, ..., n and their corresponding target values $y = \{y_i\}$, in the simplest regression task, y and x are related as

$$y = f(x) + \varepsilon \qquad \text{eq. (6)}$$

Where, the latent function f(x) is unknown and ε is often assumed to be a zero mean gaussian noise, i.e. $\varepsilon \sim N(0,\sigma^2_n)$ as shown in eq. (6). Putting GP prior over f(x) allows marginalizing it out, which means that we do need to specify its form and parameters. It makes models very flexible, since f(x) can be any non-linear function of unlimited complexity.

Thus, this work of classification and regression using Gaussian process is based on [1].

IV. EXPERIMENTAL RESULTS

Dataset: MediaEval'2013 database is used. It consists of 1000 clips each is of 45 second long taken from different locations from 1000 different songs. This database has songs distributed in 8 genres as Blues, Electronic, Classical, Country, Pop, Jazz, Folk, and Rock. Per genre consists of 125 songs. There are 53-100 unique artists per genre. Each clip is already annotated with Valence and Arousal score on a 9 point VA scale by annotators.

Setup: Visual studio is used for implementation. The experiment is done on Windows with Intel core2 dual processor, speed 2.20 GHz and RAM 1GB.
Existing system [1] only works on the dataset input i.e. it does not have facility to work on user specific input, so proposed system has facility to work on the user specific input. So, proposed system is user friendly system.
As Shown in figure 3, Y-axis is indicating time. Here, we can see existing system has taken time 42 sec to extract features. On the other hand proposed system did process of feature extraction in 34 sec

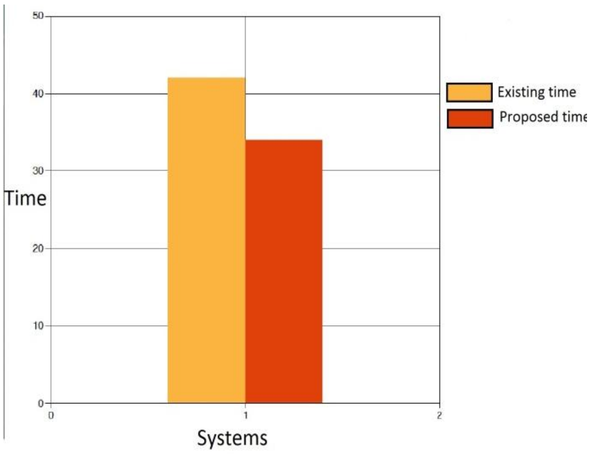

Figure 3: Feature extraction time difference

In Table 1, some of the sample clips from MediaEval dataset and from own dataset are shown with their predicted genre and emotion.

Table 1: Sample Clips and their predicted genres & emotions

Sr. No.	Song Title	Predicted Genre	Predicted Emotion
1	Sunset	Classical	Calm
2	Opening Doors	Folk	Sadness
3	Going On	Pop	Pleasure
4	Rebel Blues	Blues	Anxiety
5	O Fortuna	Classical	Surprise
6	Heat Index	Blues	Fear
7	Airtel_Horror	Rock	Anger
8	D_D_L_J	Classical	Happiness
9	Heroes	Country	Boredom
10	Jeena Yaha	Folk	Nervous

Genre classification is shown in figure 4. Here, you can see, there are 8 genre labels. On Y-axis there is a count of clips for every genre class. Highest number of clips in Classical type of genre and lowest number of clips in folk type of genre.

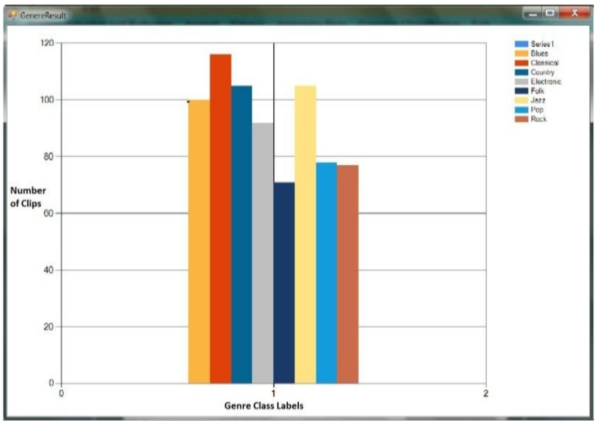

Figure 4: Clips classified into Genres

Emotion classification is shown in figure 5. There are 12 class labels of emotion. Y-axis is showing number of clips that particular class has from dataset. As shown in graph some of the emotion class like contentment, boredom and excitement don't have clips of these emotions. Less number of clips has in sadness and more number of clips in anxiety.

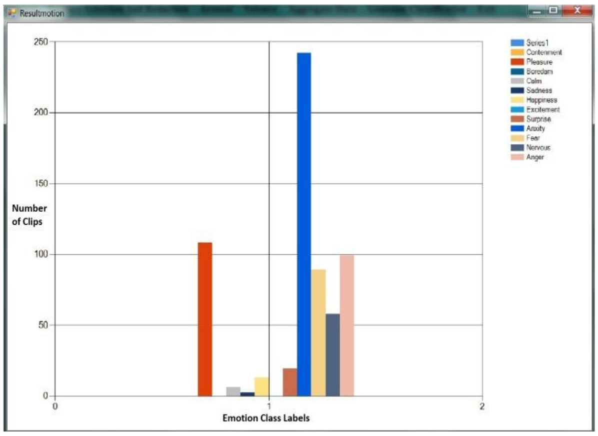

Figure 5: Clips classified into Emotion

V. CONCLUSION

Here, Gaussian Process is used for music genre classification and emotion estimation purpose. To carry out the model, for genre classification, Gaussian Process classification is utilized. Gaussian process Regression is used for emotion estimation task. Gaussian Process can be said as much more effective in the field of music information retrieval. Its predictions are truly probabilistic and also it is a non-parametric method. It is also helpful in parameter learning from training data. Study of the state of the art method i.e. support vector machines shows that Gaussian gives better results in music information retrieval specially, in emotion estimation task. We have reduced time of feature extraction using PCA. So it can be useful for making faster music information retrieval. We have applied Gaussian process on user specific clips but one can extend it to speech processing related tasks like audio recordings and voice recognition.

REFERENCES

[1] K. Markov and T. Matsui, "Music Genre and Emotion Recognition using Gaussian Process", IEEE Access, pp. 688-697, June 2014.

[2] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals", IEEE Trans. Speech Audio Process, vol. 10, no. 5, pp. 293-302, Jul. 2002.

[3] K. Chang, J.-S. Jang and C. Iliopoulos, "Music Genre Classification via Compressive Sampling", Proc. Int. Soc. Music Information Retrieval (ISMIR), pp. 387-392, 2010.

[4] M. Henaff, K. Jarrett, K. Kavukcuoglu, Y. LeCun, "Unsupervised Learning of Sparse Features for Scalable Audio Classification", Proceedings of International Society for Music Information Retrieval (ISMIR), pp. 681-686, 2011.

[5] E. Kim et al., "Music Emotion Recognition: A state of the art review", Proceedings of 11th International Society for Music Information Retrieval Conference (ISMIR), pp. 255-266, 2010.

[6] J. A. Russell, "A Circumplex Model of Effect", Journal of Personality and Social Psychology, vol. 16, no. 6, pp. 1161-1178, Dec. 1980.

[7] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval : Current directions and future challenges", Proceeding of IEEE, Vol. 96, no. 4, pp. 668-696, Apr. 2008.

[8] B. K. Baniya, J. Lee and Ze-Nian Li, "Audio feature reduction and analysis for automatic music genre classification", IEEE Int.Conf. on Systems, Man and Cybernetics, pp. 457-462, Oct. 2014.

[9] R. Typke, F. Wiering, and R. C. Veltkamp, "A survey of music information retrieval systems", Proc. Int. Conf. Music Inform. Retr., pp. 153-160 , 2005.

[10] 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.

[11] C. Rasmussen and C.Williams, "Gaussian Processes for Machine Learning", (Adaptive Computation and Machine Learning). Cambridge, MA, USA:MIT Press , 2006.

[12] Y.-H. Yang and H. H. Chen, "Machine recognition of music emotion: A review", ACM Trans. Intell. Syst. Technol., vol. 3, no. 3 , May,2012.

[13] T. Li and M. Ogihara, "Detecting emotion in music", in Proc. Int. Soc. Music Inform. Retr. Conf. (ISMIR), vol. 3., pp. 239-240, Oct., 2003.

BIOGRAPHIES

Mugdha Magare have completed Bachelor degree from S.N.J.B's K.B.J. College of Engineering, Chandwad, Nashik. Currently pursuing M.E. degree in Computer Engineering from METs Institute of Engineering, Adgoan, Nashik.

Prof. Ranjana P. Dahake working as an Associate Professor in Computer Engineering Department of METs Institute of Engineering, Adgoan, Nashik, Maharashtra, India. She has presented/published papers at national and international conferences as well as in journals on various aspects of the computer engineering. Her research areas include image processing, cloud computing and data mining.

DOI 10.17148/IJARCCE.2016.5795