

A Review on Gaussian Process Model Based Genre Classification and Emotion Estimation of Music Signals

Pranali kakad¹, Dr. Varsha Patil²

¹ ME Student, Department of Computer Engineering,,Mcoerc Nashik, Maharashtra, India

² HOD, Department of Computer Engineering,,Mcoerc Nashik, Maharashtra, India

ABSTRACT

This paper things to see methods for two tasks of Music Information Retrieval (MIR). Genre Classification and Mood Estimation are the tasks that are surveyed in this paper. Genres are the categories in which music is generally classified. It's important to understand genre categorization for efficient music retrieval system. Techniques for genre classification automate the music retrieval systems to its users. On the other hand, human generally categorize music in terms of its emotional associations. Therefore, this paper also provides review of the methods that have been proposed for music emotion recognition. This MIR tasks needs features extraction. Here, we have focused on the method Gaussian Process (GP) and its models, Gaussian Process Classification and Gaussian Process Regression for both the tasks respectively.

1. INTRODUCTION

Lots of music data is available to the users through the internet or any other sources. For users benefit, it is necessary to have efficient music information retrieval technology. Music Information Retrieval consists of tasks such as genre classification, artist identification, mood estimation, cover song identification, music annotation, melody extraction, etc. It gives efficient music search and recommendation systems, intelligent playlist generation, and other attractive applications. Here, we have considered only audio based musical signals only. As detailed work for these two tasks is possible with extracting features from audio based music signal.

Features are the characteristics of piece of music that are recorded at a particular instance of time. There are various feature types exists and their extraction algorithms too. Widely used features are as follows [1]:

- MFCC (mel frequency cepstral coefficients) – It represents short term power spectrum of a sound. It also gives amplitudes of the resulting spectrum.
- LSP (Line spectral pairs) – Speech related feature used for speech coding and for transmission over channel.
- TMBR (timbre feature) – It consists of spectral centroid, spectral flux, spectral rolloff, and zero crossings as scalar features. It shows character or quality of a musical sound, independent of pitch and loudness.
- SCF and SFM (spectral crest factor and spectral flatness measure) – These features are indicating spectral shape and used to differentiate 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 and mode.

Human categorization of music appears natural, yet it can be inconsistent, changing, and, in some cases may even seem arbitrary. Though human reactions or judgments are influenced by the audio signal, but also they are dependent on other factors such as artist fashion, dance styles, lyrics, social and political attachments, religious believes, etc. By the time new genres constantly coming into the picture while others are becoming forgotten or irrelevant. Therefore, it is difficult to come up with a commonly agreed set of music genres. Normally in music information retrieval, most popular ten types of genres are taken into consideration and those which are easily distinguishable types. Each genre classification system consists of minimum two blocks: feature extractor and classifier. Various methods for building music genre classifiers have been utilized, like, support vector machines

(SVM), compressive sampling models. In most of the cases, parametric models have been utilized. Those approaches include instances of supervised, semi-supervised and unsupervised methods.

Most of the users use genres or artist names when searching or categorizing music, but Music mostly listened because of its ability to communicate and trigger emotions in listeners. Thus, determining computationally the emotional content of music is attracting researchers towards it. Existing automatic systems for mood recognition are based on emotion representation which can be either categorical or dimensional. Categorical approaches involve finding emotional descriptors, usually adjectives, which can be arranged into groups. Given the perceptual nature of human emotion, it is difficult to come up with an intuitive and coherent set of adjectives and their specific grouping. To investigate consistent interpretation of mood categories, there are models designed to describe emotion using continuous multidimensional metrics defined on low-dimensional spaces. Most widely accepted is the Russell's two dimensional Valence-Arousal (VA) space where emotions are represented by points in the VA plane [6].

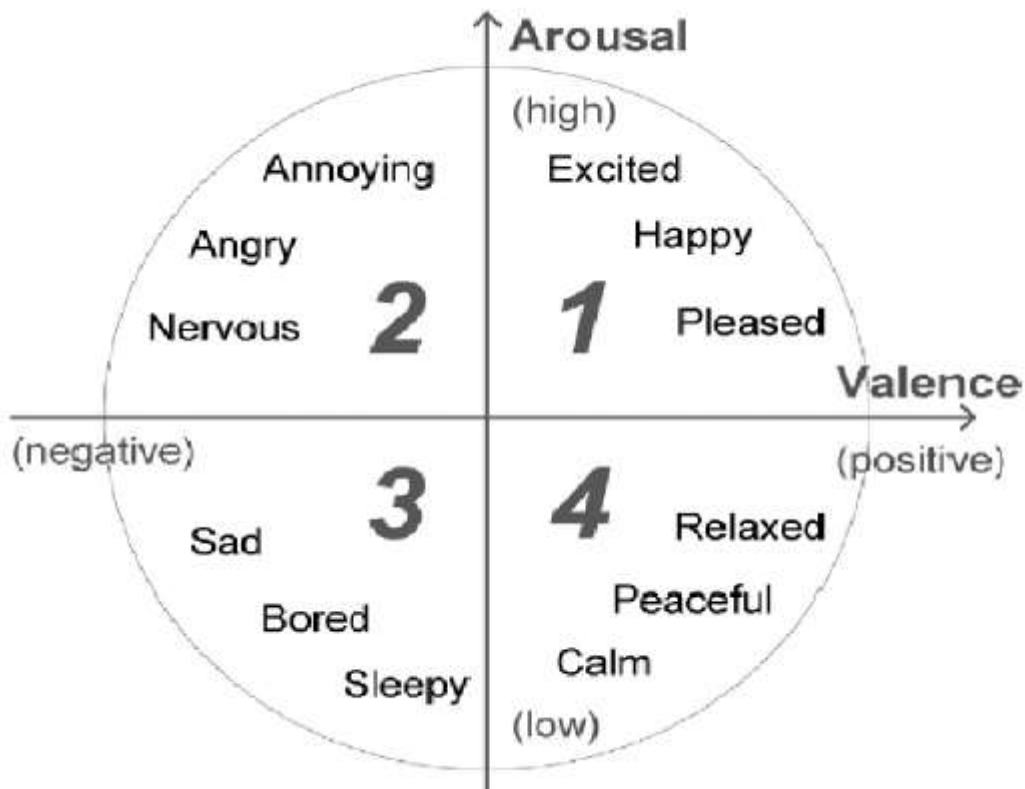


Fig -1: Two dimensional (Valence-Arousal) affective space of emotions

Figure 1 show the space where some regions are associated with distinct mood categories. In music emotion recognition, the goal is to automatically find the point in the VA plane which is matching with the emotion induced by a given music piece. Valence and Arousal are continuous and independent parameters. They can be estimated separately using the same music feature sets and different regression models.

Regression models such as Multiple Linear Regression (MLR), Support Vector Regression (SVR), as well as Multi-Level Least Squares or regression trees have been successfully applied previously to music emotion estimation.

Regression Model learning is supervised method and requires labeled training data. Finding consistent mood labels in terms of VA values is even more challenging than obtaining genre labels since emotion interpretation can be very subjective and varies among listeners. It requires music annotations by multiple experts, which is expensive, time consuming, and labor intensive. Other than area of music information retrieval, there are many other

unexplored research directions where Gaussian Process can be applied. Possibly it can be applied in speech processing and recognition where high performance is required.

2. LITERATURE SURVEY

To organize the increasingly large numbers of music files available digitally on the Web there is necessity of automatic extracting music information is important. 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 formed and used by humans for categorize and differentiate music. Musical genres have no fix definitions. It also does not have boundaries as they come up through interaction between the public, marketing, historical, and cultural factors [2].

Genre hierarchies, typically formed physically 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 music signals. In addition it provides a framework for developing and evaluating features for relating musical content. Such features can be used for other music information retrieval tasks and form the organization of most proposed music analysis techniques for music.

For representing Timbral texture, Rhythmic content and Pitch content of music signals, three feature sets were proposed and evaluated by G. Tzanetakis & P. Cook[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 located 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 music. The proposed classifier generates a dense signature to achieve a significant reduction in the dimensionality of the music signals. They also outlines the problems of content-based music information retrieval and explores 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].

The novel approach and investigated the feasibility and applicability of Gaussian Process models for genre classification and emotion estimation are proposed by Markov and Matsui [1], they investigate GPs are Bayesian nonparametric models[1].

3. SYSTEM FLOW

For genre classification, GTZAN song collection data set can be used. In this database consists of 30 second long music clips belonging to following 10 genres: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae, and Rock. MediaEval'2013 database can be used for the music emotion estimation purpose. 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

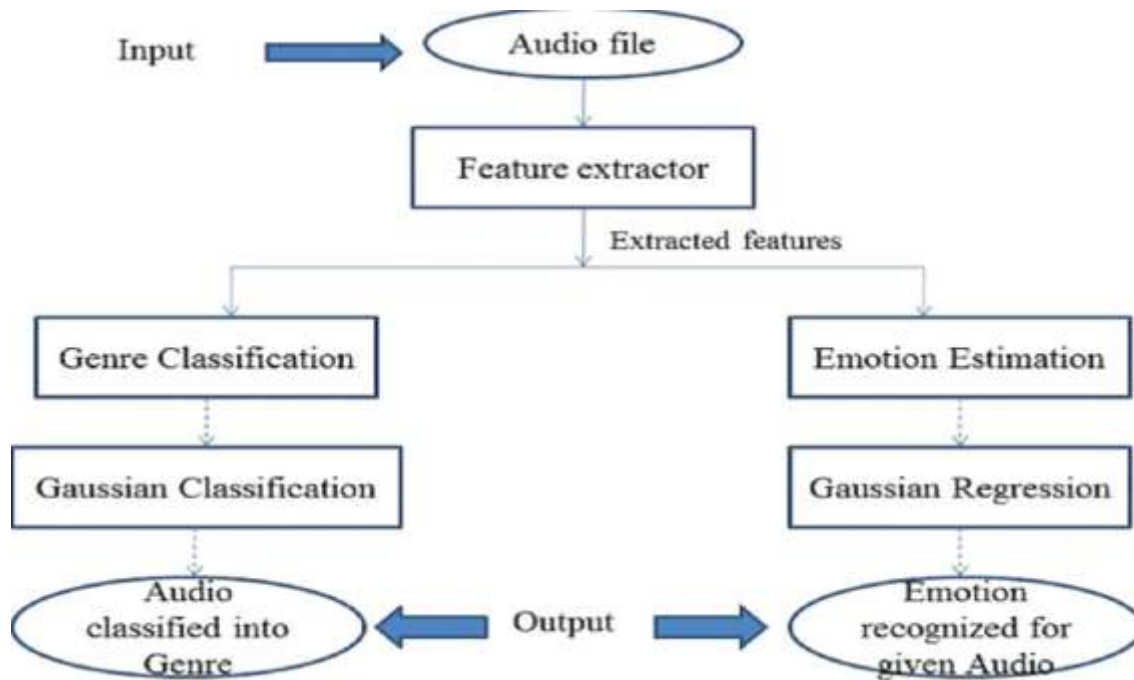


Fig -2: System Flow.

In this paper, used feature extraction methods in music signal processing studies are widely used and can be referred to as a standard set for such tasks. As shown in figure 2, both tasks can be performed after feature extraction. Thus, in this flow, first there is a feature extraction and then either of these tasks. Input is audio clip and output will be classified audio into genres and emotions recognized from input audio clips.

4. DETAILS OF GAUSSIAN PROCESS

To describe distributions over functions, Gaussian Processes (GP) are used. 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

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

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

GP can be written as,

$$f(x) \square GP(m(x), k(x, x'))$$

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

$$f \sim N(m, K)$$

where,

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

$$\begin{pmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \dots & k(x_n, x_n) \end{pmatrix}$$

Gaussian process regression is used for emotion estimation purpose and gaussian process classification is used for genre classification purpose.

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 following eq.

$$p(y | f) = \prod_{i=1}^n p(y_i | f_i) = \prod_{i=1}^n \text{sig}(y_i f_i)$$

B] Gaussian Process Regression

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

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

Where, the latent function $f(x)$ is unknown and ε is often assumed to be a zero mean gaussian noise, i.e. $\varepsilon \sim N(0, \sigma^2)$ as shown in eq. (6). Putting GP prior over $f(x)$ allows marginalizing it out, which means that we do not 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].

5 CONCLUSIONS

In this paper, we present the analysis on implementation of Gaussian Process for music genre classification and emotion estimation purpose. To carry out the model, here Gaussian Process classification is utilized for genre classification. And, Gaussian Process Regression is used for emotion estimation task. Gaussian Process is most effective in the field of music information retrieval. Truly probabilistic predictions are given by Gaussian Process and also it is a nonparametric method. It is also useful in parameter learning from training data. In compare with state of the art techniques, gives better results in music information retrieval. Gaussian may also be useful in speech recognition area.

6. 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.