ISSN No: 2454-9614

Evaluating the Multi-Emotion Similarity Preserving Embedding Feature Mining for Music Emotion Analysis

V. Jayasudha, R. Vijaysai*

Department of computer Science and engineering, K.S.Rangasamy College of Technology,
Namakkal, Tamil Nadu, India.

*Corresponding Author: V.Jayasudha

E-mail: vijaysair@ksrct.ac.in

Received: 11/11/2015, Revised: 13/12/2015 and Accepted: 12/03/2016

Abstract

Music is often referred to as "language of love", and it is natural for us to categorize music in relation to their emotional associations. Countless features, such as harmony, timbre, interpretation and lyrics emotion influence, and also change the mood of a piece can have their duration. Human hypotheses are based on real experiences and proven by psychological paradigms on people. The psychology of the first hypothesis between music and emotion on the basis of intuition or experience and then, human experiments to validate the hypothesis. The existing technology to a systematic and quantitative framework. Formulate the task as a multi-label similarity. Proposals in this survey, a novel multi-label dimensionality reduction algorithm synchronized Multi-Emotion similarity preserving embedding (ME-SPE), to assign the original high-dimensional representations in a low-dimensional function Unterraum, in the hope that a clearer link between the functions and emotions could be detected. The performance of the state-of-the-art approaches can be further improved in most of the criteria.

Reviewed by* **ICETSET'16 *organizing committee*

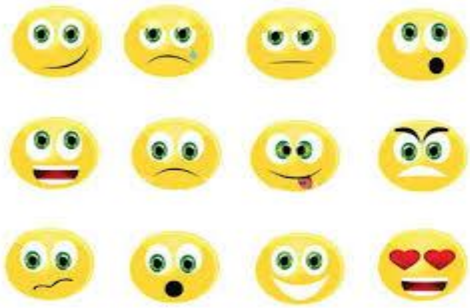
Keywords*: Music emotion analysis, feature mining, multi-label dimensionality reduction*

1. Introduction

Music is an art from whose medium is sound. The word derives from Greek møusikh´ (te´cnh), “(art) of the Muses”. Nowadays, music is everywhere. From small television commercials, music videos and shopping centers to the more traditional music sell (albums, both on physical or digital format, and tickets), radio play or live events (concerts, gigs, festivals, etc), cannot help consider now it as an industry.

Music has always been transformed by advances in technology. Examples of technologies that transformed the way music was produced, distributed and consumed include musical instruments, music notation, recording and more recently digital music storage and distribution. Recently portable digital music players have become a familiar sight and online music sales have been steadily increasing. It is likely that in the near future anyone will be able to access digitally all of recorded music in human history. In order to efficiently interact with the rapidly growing collections of digitally available music it is necessary to develop tools that have some understanding of the actual musical content.

As defined in Drever’s psychology dictionary emotion is “a mental state of excitement or perturbation, marked by a strong feeling, and usually an impulse towards a definite form of behaviour”. In order to understand such a complex concept of emotion and distinguish it from other psychological states such as reflex, motive, and attitude, Ekman presented a classical categorical model of emotions based on the human facial expressions, which divides emotion into six basic classes: anger, happiness, surprise, disgust, sadness, and fear.



Being able to extract various features of the music from the audio signal, or descriptors, is a key to the creation of automatic content description tools which would be highly useful for music analysis, retrieval and recommendation.

Such descriptors are often classified as either low, mid or high-level descriptors, depending on the degree of abstraction of the descriptor [Lesaffre **05**]. Features denoted as low-level are usually those which are closely related to the raw audio signal, computed from the signal in either a direct or derived way. Such descriptors will usually not be very musically meaningful for end-users, but are of great value for computational systems. Examples of low level descriptors are Harmonic Pitch Class Profile (HPCP).

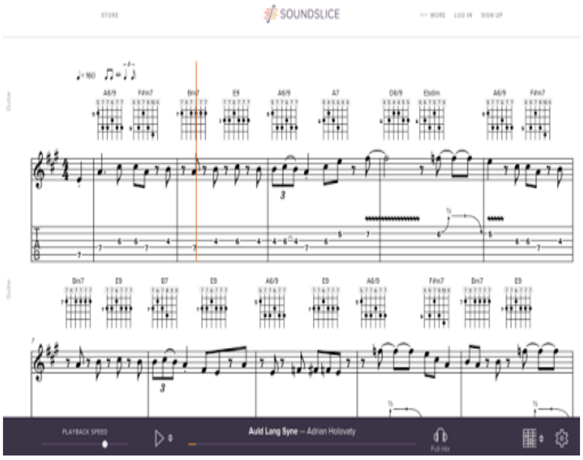


Figure 1.1 the music information plane

Mid to high-level descriptors are those we would consider more musically meaningful to a human listener rather than a machine. Examples of such descriptors are the beat, chords, melody, and even more abstract concepts such as mood, emotions, or the expectations induced in a human listener by a piece of music.

When discussing music description, Serra proposes the \Music Information Plane" [Serra **05]**, a plane where the relevant information about music is placed in the context of two dimensions; one is the abstraction level of the descriptors (from physical to knowledge levels) and the other includes the different media (audio, text, image).

2. Related Work

Min-Ling Zhang et al.[12] reviewed the multi label learning. Multi-label learning studies the problem where each example is represented by a *single* instance while associated with a set of labels simultaneously. During the past decade, significant amount of progresses have been made towards this emerging machine learning paradigm. Firstly, fundamentals on multi-label learning including formal definition and evaluation metrics are given. Secondly and primarily, eight representative multi-label learning algorithms are scrutinized under common notations with relevant analyses and discussions. Thirdly, several related learning settings are briefly summarized. Traditional supervised learning is one of the mostly-studied machine learning paradigms, where each real-world object (example) is represented by a single instance (feature vector) and associated with a single label. Although traditional supervised learning is prevailing and successful, there are many learning tasks where the above simplifying assumption does not fit well, as real-world objects might be complicated and have multiple semantic meanings simultaneously.

X. Wang et al. [1] stated that music recommendation systems rely on collaborative filtering or content-based technologies to satisfy users' long-term music playing needs. Given the popularity of mobile music devices with rich sensing and wireless communication capabilities, a novel approach to employ contextual information collected with mobile devices for satisfying users' short-term music playing needs. Present a probabilistic model to integrate contextual information with music content analysis to offer music recommendation for daily activities, and present a prototype implementation of the model. Considerable attention has focused recently on context-aware music recommender systems (CAMRSs) in order to utilize contextual information and better satisfy users' short-term needs. The model uses a Bayesian framework to seamlessly integrate context-aware activity classification and music content analysis.

M. Muller et al. [2] stated that Music signal processing may appear to be the junior relation of the large and mature field of speech signal processing, not least because many techniques and representations originally developed for speech have been applied to music, often with good results. This paper provides an overview of some signal analysis techniques that specifically address musical dimensions such as melody, harmony, rhythm, and timbre. The goal is to demonstrate that, to be successful, music audio signal processing techniques must be informed by a deep and thorough insight into the nature of music itself. MUSIC is a ubiquitous and vital part of the lives of billions of people worldwide. Musical creations and performances are among the most complex and intricate of our cultural artifacts, and the emotional power of music can touch us in surprising and profound ways. This paper concerns the application of signal processing techniques to music signals, in particular to the problems of analyzing an existing music signal (such as piece in a collection) to extract a wide variety of information and descriptions that may be important for different kinds of applications. That there is a distinct body of techniques and representations that are molded by the particular properties of music audio such as the pre-eminence of distinct fundamental periodicities (pitches), the preponderance of overlapping sound sources in musical ensembles (polyphony), the variety of source characteristics (timbres), and the regular hierarchy of temporal structures (beats). These tools are more or less unlike those encountered in other areas of signal processing, even closely related fields such as speech signal processing.

Y.-H. Yang et al. [5] stated that the proliferation of MP3 players and the exploding amount of digital music content call for novel ways of music organization and retrieval to meet the ever-increasing demand for easy and effective information access. This article provides a comprehensive review of the methods that have been proposed for music emotion recognition. Moreover, as music emotion recognition is still in its infancy, there are many open issues. Review the solutions that have been proposed to address these issues and conclude with suggestions for further research. The way that music information is organized and retrieved has to evolve in order to meet the ever-increasing demand for easy and effective information access. According to a study of social tagging on Last.fm1, a popular commercial music website, emotion tag is the third most frequent type of tag (second to genre and locale) assigned to music pieces by online users. Emotion based music retrieval has received increasing attention in both academia and the industry.

E. Coutinho et al. [3] stated that there is strong evidence of shared acoustic profiles common to the expression of emotions in music and speech, yet relatively limited understanding of the specific psychoacoustic features involved. This study combined a controlled experiment and computational modeling to investigate the perceptual codes associated with the expression of emotion in the acoustic domain. The empirical stage of the study provided continuous human ratings of emotions perceived in excerpts of film music and natural speech samples. The computational stage created a computer model that retrieves the relevant information from the acoustic stimuli and makes predictions about the emotional expressiveness of speech and music close to the responses of human subjects. Show that a significant part of the listeners' second-by-second reported emotions to music and speech prosody can be predicted from a set of seven psychoacoustic features: loudness, tempo/speech rate, melody/ prosody contour, spectral centroid, spectral flux, sharpness, and roughness. The implications of these results are discussed in the context of cross-modal similarities in the communication of emotion in the acoustic domain.

3. Overview of the Project

In musical emotion analysis, a song may convey or evoke more than one emotions. Some researchers therefore formulate it as a multi-label learning problem. Unlike single-label learning in which each data point belongs to only one category, multi label learning is more general than the single-label case that each data point might be associated with multiple labels. More importantly, an implicit assumption in single-label learning is that the labels are mutually exclusive while in multi-label learning it is possible that the labels are correlated with each other. Driven by various applications such as image classification and text categorization, many multi-label learning algorithms have been proposed, such as multi-label *k*-nearest neighbour, multi label support vector machines, multi-label neural networks, etc. A more comprehensive review on multi-label learning algorithms could be found in. Furthermore, the music signals often have a huge number of features which may contain a large amount of redundant information and thus cause the high computational cost and poor performance of the analysis task. Therefore, multi-label dimensionality reduction becomes our first choice for the task of music emotion analysis. To the best of our knowledge, there is no work on multi-label dimensionality reduction for music emotion analysis. However, multi-label dimensionality reduction itself is already an active research area in machine learning. Yu et al. proposed a method called multi-label informed latent semantic indexing to preserve the information of data and meanwhile capture the correlations between the multiple labels. Arenas-Garca et al. presented the sparse kernel ortho normalized partial least squares to handle the multi-label data. Sun et al. proposed the hyper graph spectral learning, which generalize the graph Laplacian to the hyper graph Laplacian for multi label applications. Park and Lee extended the traditional linear discriminant analysis to the multi label version by applying the copy transformation. Wang et al. proposed another multi-label linear discriminant analysis algorithm by taking advantage of label correlations. Zhang and Zhou introduced a multi-label dimensionality reduction algorithm by maximizing the dependence between data and corresponding labels. Ji et al. proposed a shared subspace learning model for multi-label classification. Other well-known dimensionality reduction schemes, such as nonnegative factorization, canonical correlation analysis, and sparse coding, have also been extended for multi-label classification.

3.1 Existing system

Existing works in psychology first made hypothesis between music and emotion based on the intuition or experience and then conducted human experiments to validate the hypothesis. Different with them, the proposed technique intends to provide a systematically and quantitative framework. We formulate the task as a multi-label dimensionality reduction problem according to the following two considerations. First, dimensionality reduction technology targets to find the intrinsic structure embedded in the original feature space of the raw data for the given goal. To discover the relationship between the music and emotion, the raw data is the music and the original feature space can be defined as the music signal representation. The given goal can be defined as the known human's emotion response to some music. Second, a song may evoke more than one emotion, so we explore the dimensionality reduction under the framework of multi-label learning

3.1.1 Issues in Existing system

Single-label learning in which each data point belongs to only one category. Multi label learning is more general than the single-label case that each data point might be associated with multiple labels. More importantly an implicit assumption in single-label learning. Labels are mutually exclusive while in multi-label learning it is possible that the labels are correlated with each other. The emotions cannot be finding perfectly. It is only finding for foreign instrumental music. In Low level music the emotions cannot be finding.

3.2 Proposed system

Multi emotion similarity between conservation is used to the pitch, rhythm, and sound like multi emotion tone. Multi-label classification is used to classify the emotion such as happy, sorrow, hunger, etc. short-term Fourier transmission is used to digital signal and data processing of the sound and trained from the music signal representation, notice, that after engages pin, the signals are of course represented by second-order tensors (i.e. , matrices). Vectors zing these matrices before trend spotting not only increases the computational cost dramatically, but also destroys the frequency-time structure of the data. In order to adapt it to the second-order signals, we expand ME-SPE on its bilinear version.

Algorithm

- Short time Fourier transforms (STFT).
- Multi-label classification

Advantage

Also lead to general emotion classification, for example, the distinction between "angry", "wonder" and "happy" from "relax", "sad" and "quiet". For all types of music, song and voice find the emotions perfectly.

1. Pitch Detector Module

Musical event, note, as well as higher level music structures: rhythm, harmony, and melody. Note is the atom of music, rhythm, harmony, and melody in music be described by three main characteristics: duration, pitch and amplitude. Musical duration concerns the length of a sound in time, and is represented by different note value symbols, e.g., with/without a flag. Pitch is the perceived fundamental frequency of the note, indicated by the different vertical positions on the five-line staff. Amplitude is used to identify the degree of loudness of a note. Rhythm, one of the most important factors affecting emotional expression, generates a regular, repeated pattern of sounds or movements by combining notes of different durations and amplitudes. Harmony, another element related to emotion expression in music, involves the vertical aspect of pitch and how various sounds go together to create pleasing and interesting combinations. Melody is the horizontal unfolding of pitch over time and creates shapes that we identify as songs. Its ascent and descent may be associated with different emotions. Some other components of music, such as loudness and mode, may also contribute to conveying emotions.

2. Neural classifier trained model

A large number of feature sets have been proposed to represent music audio signals. Typically they are based on some kinds of time-frequency representations, e.g. the short time Fourier transform (STFT). Mel-frequency cepstral coefficients (MFCC) are perceptually motivated features that are also based on the STFT. The calculation of MFCC is based on the following steps:

1). Computation of the mel-warped spectrum with a bank of overlapping band-pass filters;

2). Taking the logarithm of the magnitude of each resulting band; and

3). Calculating the Discrete Cosine Transform (DCT) on the resulting bands.

The rhythm feature set is based on detecting the most salient periodicities of the music signal. It is extracted by deriving a so-called beat histogram. Specifically, the enhanced autocorrelation function of the temporal envelop of the musical signal is calculated and its dominant peaks, which correspond to various periodicities of the signal, are detected and accumulated into a histogram, where each bin corresponds to the beat period in beats-per-minute.

3. Analyze the music data

The **music** is now extracted from feature extractor is passed in matcher system and this matcher system matches the trained sounds with recorded sound. the feature sets on EMOTIONS and CAL-500 datasets are quite different. For the 72- dimension features in EMOTIONS dataset, we use them because these features have been reported to be effective for music emotion recognition. In addition to these widely used music features, we also want to explore the music features from the very original feature space. We have observed that many music signal features, such as spectral centroid, spectral roll off, spectral flux, and MFCC, are calculated based on the magnitude of STFT on the time-frequency diagram. Therefore, we utilized the normalized magnitude spectrogram as the features of the data in the CAL-500 dataset.

4. Emotion analysis

Pre-processing, Framing, Windowing, FFT & Mel filter bank and Frequency wrapping processes of MEDC feature extraction are same as MFCC feature extraction. in music signals that convey or evoke emotions, we propose a novel multi-label dimensionality reduction algorithm dubbed Multi-Emotion Similarity Preserving Embedding (ME-SPE) to map the original high-dimensional representations into a low-dimensional feature subspace, in which we hope that a clearer linkage between the features and emotions could be discovered. Let X = R*D* be the high-dimensional feature space of the music signal, and there is an emotion set E including *m* emotion labels. The emotions associated with the music **x** constitute a subset of E, which can be represented as an *m*-dimensional binary vector **y**, with 1 indicating that the music is able to convey the corresponding emotion

4. Conclusion & Future Work

To discover the relationship between music and emotion via computational approach. A novel multi-label dimensionality reduction technology named ME-SPE as well as its bilinear extension BME-SPE are proposed. The three questions proposed in the Abstract have been answered as follows. First, we find that the mean and standard deviation of spectral flux, the first component of MFCC, and the first and the second beat histogram peaks in bpm are the most important features that essentially convey emotions. Second, the extent that these features contribute to representing emotions is calculated. Third, the findings in our work are consistent with some of the observations from psychology. Specifically, the mean and standard deviation of spectral flux reflect how fast the pitch of a song changes and how inconsistent the change is.

These features can trigger the human brain to extract higher level music features such as harmony and melody, which matches the argues in psychology that harmony and melody features are closely related to emotion responses. The first MFCC coefficient is an overall measure of the signal loudness, which supports the claim that the loudness of the music contributes to conveying emotions.

The periods of the first and the second peaks in bpm reflect the rhythm of the music, which is consistent with the observations that rhythm features are important in emotion expressions. For the future work, we are especially interested in the second order low-dimensional feature space generated by BME-SPE, which achieves good performance on CAL-500 dataset. We will study the physical x meaning of these features, and their relations with the findings from psychology. Moreover, we are going to explore the weighted extension of ME-SPE and BME-SPE, which offers the opportunity to model the emotion extent (not just yes or no) of the music.

References

[1] X. Wang, D. Rosenblum, and Y. Wang, “Context-aware mobile music recommendation for daily activities,” in Proc. 20th ACM Multimedia, 2012, pp. 99–108.

[2] M. Muller, D. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

[3] E. Coutinho and N. Dibben, “Psychoacoustic cues to emotion in speech prosody and music,” Cognition and Emotion, vol. 27, no. 4, pp. 658–684, 2013.

[4] A. Wieczorkowska, P. Synak, and Z. Ra´s, “Multi-Label Classification of Emotions in Music,” in Intelligent Information Processing and Web Mining, ser. Advances in Intelligent and Soft Computing, M. Klopotek, S. Wierzchon, and K. Trojanowski, Eds. Springer Berlin/Heidelberg, 2006, vol. 35, pp. 307–315.

[5] Y.-H. Yang and H. H. Chen, “Machine recognition of music emotion: A review,” ACM Trans. Intell. Syst. Technol., vol. 3, no. 3, pp. 40:1–40:30, May 2012.

[6] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen, “A Regression Approach to Music Emotion Recognition,” IEEE Trans. Audio, Speech, Language Process., vol. 16, no. 2, pp. 448–457, 2008.

[7] Y.-H. Yang, C.-C. Liu, and H. H. Chen, “Music emotion classification: A fuzzy approach,” in Proc. 14th ACM Multimedia, 2006, pp. 81–84.

[8] K. Yu, S. Yu, and V. Tresp, “Multi-label informed latent semantic indexing,” in Proc. 28th SIGIR, 2005, pp. 258–265.

[9] M. Zentner, D. Grandjean, and K. R. Scherer, “Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement,” Emotion, vol. 8, no. 4, pp. 494–521, 2008.

[10] M.-L. Zhang and Z.-H. Zhou, “Multilabel neural networks with applications to functional genomics and text categorization,” IEEE Trans. Knowl. Dat Eng., vol. 18, no. 10, pp. 133 1351, 2006.

[11] 8 “Ml-knn: A lazy learning approach to multi-label learning,” Pattern Recogn., vol. 40, no. 7, pp. 2038–2048, 2007.

[12] A review on multi-label learning algorithms,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 8, pp. 1819–1837, 2014.

[13] Y. Zhang and Z.-H. Zhou, “Multilabel dimensionality reduction via dependence maximization,” ACM Trans. Knowl. Discov. Data, vol. 4, no. 3, pp. 14:1–14:21, 2010.

[14] J. Arenas-Garca, K. Petersen, and L. Hansen, “Sparse kernel orthonormalized pls for feature extraction in large data sets,” in NIPS 19, 2007, pp. 33–40.

[15] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” Neural Comput.,vol. 15, no. 6, pp. 1373–1396, 2003.