2009 Second International Symposium on Computational Intelligence and Design

Automatic Music Emotion Classification Using a New Classification Algorithm

Xiaoyu Sun
College of Computer Science
Zhejiang University
Hangzhou 310027, PR. China
sunxiaoyufish@gmail.com

Yongchuan Tang
College of Computer Science
Zhejiang University
Hangzhou 310027, PR. China
tyongchuan@gmail.com

***Abstract*—Music emotion is a special emotion that is aroused by music which is a media that can convey human affection. Music emotion classification is a popular topic in recent years. The mood of a music clip describes emotional expression. It is helpful in music understanding, music retrieval and some other interesting music related application. In this paper, a method is proposed using a framework named information cell mixture models (ICMM) to automate the task of music emotion classification. This framework has potential application in both unsupervised concept learning and supervised classification learning. This framework is acceptable for music mood classification because emotion is a vague concept and has a cognitive structure. The application of ICMM is also suitable for music emotion classification.**

Keywords-music emotion; classification; information cell; mixture models;

I. INTRODUCTION

Music is popular in people leisure time. At the same time, there is more and more music on the internet and people's computers. It is time-consuming to organize all music or to search a special music by manual work. How to facilitate music organization and searching automatically is a potential work.

Music emotion plays critical role in many applications for music. For example, when some young people have a party on week ends, they may want to choose an exciting music; while when an individual wants to relax himself at home, he may choose a calm music. The mood information have crucial applicability in these scenes. Music play-list generation, music piece searching and music management based on affective tags could propose more convenient way to achieve a piece for a particular application.

Two critical works for music emotion tags marking are feature extracting and music classification. This paper using a framework for audio analysis named MARSYAS [1] to achieve audio features and using Information Cell Mixture Models (ICMM) [2] to classify music in library. It is a new attempt to estimate the mood inferred in each music automatically by ICMM and this method is firstly proposed to apply to emotion classification.

A.. Related Works

Most previous works on sentiment detection which applied standard pattern recognition procedure and data mining procedure employed traditional machine learning algorithms. In [3] , a model based on Bayesian networks was proposed to automatically recognize the expressive content of MIDI piano improvisations. In [4], the classifier is a BP neural network. In [5], the support vector machine (SVM) is trained using the extracted features.
Some other attempts tried to predict the emotion that the music is trying to convey such as [6], a hierarchical framework was proposed to detect the mood in a music clip by means of intensity, timbre, and rhythm features. In [7], system-identification techniques are used to create the models.

B. Issues in Music Mood

In order to build an automatic mood detection system from acoustic music data, there are several issues that need to be considered beforehand.
The first issue is there is debate over whether music emotion detection is feasible.

Different individuals have different perceptions over one same clip. It is argued that music mood is subjective and it is meaningless to detect music affective. However, within a given cultural context, there are agreements about the mood inspired by the same music clip [8].

The second consideration is how to describe mood detected from music. Some approaches use adjective such as gloomy, hopeful, cheerful, dreamy. Hevner has proposed a checklist [9] composed of 67 adjective and clustered them to 8 classes. However, the adjective checklist is different among different research context and immense. In this paper Thayer's dimensional approach is applied. Russell proposed a mood model on two bipolar dimensions [10] instead of some adjectives. Thayer adapted this model to music [11]. This model is composed of two dimensions, one is energy and the other is stress. In affective computing, they are activity and valence. The energy dimension corresponds to arousal or activity while the stress dimension corresponds to pleasure or valence. Based on the level of energy and stress, Thayer's model divides music mood into four clusters: contentment, depression, exuberance and anxious. The four basic mood clusters

978-0-7695-3865-5/09 $26.00 © 2009 IEEE
DOI 10.1109/ISCID.2009.281

discriminate mood respond to music explicitly. In out scheme, music database is indexed on four labels of music mood: contentment, depression, exuberance and anxious.

II. Feature Extraction

Marsyars is a flexible framework for building and integrating audio analysis tools. In this work, Marsyars is used to extract musci clip features. The currently supported featrures are: spectral centroid, spectral moments, spectral flux, pitch, harmonicity, mel-frequency cepstral coefficients (MFCC), linear prediction reflection coefficients.

A number of techniques for audio analysis have been proposed. A robust multi-feature music/speech discriminator is described in [12]. A retrieval-by-similarity system for isolated sounds has been developed at Muscle Fish LLC [13]. However,Marsyas system is more flexible to customize music features user wanted and it supplies a practical user interface.

In this paper, we have used a number of features that have been implemented in Marsyas system. These features consiste of spectral controid, spectral moments, spectral flux, pitch, harmonicity, mel-frequency cepstral coefficients, linear prediction reflection coefficients, zero crossings, rms, spectral rolloff and others.

There are 100 clips in our library, they are from 15 west classical musics, the featrues number is 70.

III. Mood Classification

Information cell mixture models [2] is proposed for vague concept modeling and learning where an information cell mixture model(ICMM) is actually a vague concept cell and is represented by a semantic cognitive structure comprising of a prototype set, a distance function and a density function.

An unsupervised information cell formation algorithm is introduced to learn an ICMM using positive neighborhood function if only positive data set is available for a concept. A supervised information cell formation algorithm is also developed to learn an ICMM using both positive-neighborhood and negative neighborhood functions if both positive and negative data sets are available for a concept [2]. In this paper, music emotion space is considered a complete concept therefore an unsupervised information cell formation algorithm is adopted.

LA = {L1,..., Ln } is a set of labels (words) for elements from Rm. For each label Li, we assume that it can be interpreted as 'similar to Pi ', where Pi ∈ Rm is a set of prototypical cases of concept Li.

An information cell for vague concept 'Li = similar to Pi ' on the domain Rm is represented by a ternary semantic structure <Pi, di, δi> , where Pi is a set of prototypes of concept Li and in this paper Pi is a single element in Rm. di is a distance function on Rm ×Rm such that for any X, Y Rm, d(X, Y) = (X − Y)T(X − Y), the density function δi(ε) is an additive normal density function introduced firstly in [2] as follows:

δ (ε| ci , σi) = f (ε| ci , σi) + f (ε| −ci , σi)

where f (ε| ci , σi) is a normal density function.

For any Li ∈ LA and ε≥ 0 the ε-positive neighborhood NNLi of cell Li is defined as follows:

$$NN\varepsilon Li = \{X : d(Xi,P) \le \varepsilon\} \qquad (1)$$

The positive-neighborhood function of information cell Li is given by:

$$\mu+Li(X)=\delta i(\{\varepsilon: X\in NN\varepsilon Li\}) = \int_{d(X,Pi)}^{+\infty} \delta i(\varepsilon)d\varepsilon \qquad (2)$$

An information cell mixture model (ICMM) is formal represented as LP=< LA, Pr> , where LA is a set of information cells Li, Li = <Pi, di, δi> for i = 1,...,n and Pr = {ri : i=1,…n} is a probability distribution on LA such that $\sum_{i=1}^{n} ri = 1$.

The music emotion space is considered a complex concept where a set of information cells to represent the variant emotion and each information cell Li is assigned a probability ri. The music emotion space classification work is equivalent to learn an ICMM from a data set DB, where DB is considered as the set of positive instances of ICMM. For learning an ICMM LP from a data set DB an objective function J(LP | DB) is introduced. Then a learning algorithm for optimizing the objective function J(LP | DB) is proposed. Details of the unsupervised information cell formation algorithm can be find in reference [2].

IV. Conclusion

The approach detailed in this paper proposed a attempt of employing a new classification algorithm ICMMs to the emotion classification and index. The core trait of ICMMs is that it is based on a ternary semantic strcucture of vague concept and it is very efficient for high dimentional concept modelling. The features extracted by the Marsyars system is a high dimentional features set, it is a motivation of the applying of this algorithm for the sentimental detection.

Acknowledgements

This work is sponsored by the National Natural Science Foundation of China (NSFC) under Grant No. 60776798.

References

[1] George Tzanetakis, "MARSYAS: A framework for audio analysis," http://marsyas.sness.net.

[2] Yongchuan Tang and Jonathan Lawry, "ICMMs: information cell mixture models for concept modeling and learning," submitted.
[3] Mion, "Application of Bayesian networks to automatic recognition of expressive content of piano improvisations," Proc. Stockholm Music Acoustics Conf. (SMAC'03), 2003, pp.557–560.
[4] Yazong Feng and Yueting Zhuang, "Popular music retrieval by detecting mood," SIGIR'03, 2003, pp. 375-376.
[5] WANG Muyuan and Zhang Naiyao, "Emotion recognition of western tonal music using support vector machine," Chinese Journal of Electronics, 2006.
[6] Lu, L., Liu, and D., Zhang, H.J., "Automatic mood detection and tracking of music audio signals," IEEE Trans. on audio, speech and language processing 14(1), 2006, pp. 5–18.
[7] Mark D. Korhonen and David A. Clausi, "Modeling Emotional Content of Music Using System Identification," IEEE Trans. on systems, man, and cybernetics, 2006, vol.36, NO.3, pp. 588-599.
[8] E. Radocy and J. D. Boyle, "Psychological foundations of musical behavior," Springfield, IL: Charles C. Thomas, 1988.
[9] K. Hevner, "Expression in Music: a discussion of experimental studies and theories," Psychol. Rev., 1935, vol.42, pp.186-204.
[10] J.S.Russell, "A circumplex model of affect," J. Personality Social Psychology, 1980, vol.39, pp. 1161-1178.
[11] R.E.Thayer, "The biopsychology of mood and arousal," Oxford Univ. Press, 1989.
[12] Scheirer E. and Slaney M.,"Construction and evaluation of a robust multifeature speech/music discriminator," IEEE Trans. on acoustics, speech and signal processing (ICASSP'97), 1997, pp. 1331-1334.
[13] Wold E,"Content-based classification, search and retrieval of audio," IEEE Multi-media, 1996, pp. 27-36.