A Framework for Multi-Label Music Mood Classification

Ei Ei Pe Myint
University of Computer Studies, Yangon
eieipemyint@gmail.com

Abstract

Music is the effective communication medium among people. Studying music mood can help in music understanding, music retrieval, and some other music-related applications. This paper presents a hierarchical framework with a new mood taxonomy model to automate the task of mood classification from acoustic music data based on western music psychology theory. This system proposes hierarchical framework with new mood taxonomy model. The proposed mood taxonomy model is combined by the Thayer's 2 Dimension model and Schubert's updated Hevner adjective checklist. The 60 famous English songs are used as the standard database in this system which is created by literature. The verse and chorus part from the whole song is extracted manually for processing in this proposed system. The extracted music clip is segmented by image region growing method to separate homogenous part on the entire music clip. Then, the feature sets from the separated music trimmed are extracted to inject the Fuzzy Support Vector Machine (SFVM). To solve the multi-label classification problem, one-against-one (O-A-O) multi class classification method are used. The hierarchical framework with new mood taxonomy model has the advantage of reducing the number of classifier used for O-A-O approach.

1. Introduction

Music is very important for our everyday life. It is not only for entertainment or pleasure. It connects people, joins lives, experiences, emotions and hearts. Because of the booming of communication technology, there is more and more music on personal computer, in music libraries and on the Internet. In order to facilitate music management such as music search, music information retrieval and automated play-list generation need to be created for each music piece. Although the traditional information such as the artist, the album, or the title of music work remains important, these tags have limited applicability in many music related applications. Since the preeminent functions of music are social and psychological, the most useful characterization would be based on four types of information: the style, genre, similarity and emotion. Music classification and retrieval by perceived emotion is natural and functionally powerful since people can share their emotion among each other by using music as a communication medium. However, to the best our knowledge few system claims to be able to automatically retrieval music by mood because of having two obstacles lies on this approach: one is that there is no computational mood model yet, and the other is that mood is an item related with cultural background and involved scenario.

Though the relationship between music and perceived emotion has been studied by psychologist for decades, the boom of music emotion classification/recognition can be dated back to within 10 years [1-9]. Past approaches towards automated classification of emotions in music modeled the learning problem as a single-label classification [1-4], regression [5, 6], or multi-label classification [7-9] task. Actually, music may evoke more than one different emotion at the same time. Single label and regression cannot model this multiplicity. Therefore this paper focuses on multi-label classification methods.

All the music mood classification methods [1-9] use some music mood taxonomy based music psychological theory of western culture. Most of the single label classification followed Thayer's 2D model [1-6] since multi-label classification uses Fransworth adjective model [7-9].

The primary contribution of this paper is proposing a new mood taxonomy model by combining Thayer's 2D model and Schubert's updated Hevner adjective checklist [10]. The aggregation of hierarchical framework with our new mood taxonomy model can reduce the number of classifier use while it processing.

The second contribution is mood tracking system which is followed by music mood classification. According to the literature, the mood may well change one or more times within a single piece. Therefore mood change in entire music is tracked before it is classified in this system. It is adopted by image region growing method for extracting continuous song segment which contained homogenous mood.

Support vector machines (SVMs) are learning machine which are intended to binary classification problems. In order to increase the applicability of SVMs, it is necessary to extend them to multi-classification. To solve multi class classification problem, Knerr, et al. [11] proposed one against one SVM (O-A-O SVM) model. To reduce outliers effectively, Fuzzy membership function is combined to each input point and reformulates SVM in this system, such that different input points can make different contributions to the learning of decision surface proposed by Lin et al. [12].

2. Related Work

One common objection to music mood detection is that the emotional expression and perception of music is subjective and it depends on many factors including culture, education, and personal experience. Thus, for the same music piece, different musicians may have different performances, while different individuals might have different perceptions. Therefore, it is usually argued that the mood is too subjective to be detected. However, as much research such as [13] has shown, musical sounds, with certain patterns or structures, usually have inherent emotional expression. Moreover, Juslin [14] indicated that these emotions are able to be communicated. It was also found that, within a given cultural context, there are major agreements among individuals regarding the mood elicited by music. Therefore, it is possible to build a mood classification system in a certain context.

In [1-9] presented the music mood/emotion classification framework by using various feature sets, mood taxonomy models and classification methods. Most of them used audio feature analysis tools, PsySound2 [15] and Mersyas [16] for feature extraction. All the music mood classification used 20 to 30 second music trimmed for each of their approaches. Before classification, those all system first down sampled their dataset into 16 to 22.05 kHz, mono channel.

Thayer's model of mood is adopted as the basic of mood taxonomy for mood detection in Liu *et al*. [1] and Yang *et al*. [4]. However, most of the multi-label music mood/emotion model except [9] adopted Fransworth's group of adjectives as their basic model.

Previous single label classification framework used GMM, BP Neural Network, Bayesian network, Fuzzy and SVM classifiers [1-6] for music mood classification, though multi-label music mood classification used SVM and k-NN classifiers.

3. Mood Taxonomy

Most of the single label classification such as [1-6], the emotion classes are defined in terms of arousal/energetic (how exciting or calming) and valance/stress (how positive or negative). For example, the emotion class can be divided into the four quadrants known as exuberance, anxious/frantic, contentment and depression in Thayer's arousal and valance model. In multi-label classification frame work which allows assigning more than one emotion class to the same song in response to the fact that human perception of music emotion is not uniform. Most of them adopted Fransworth's group of adjectives as their basic mood taxonomy model. It consists of thirteen adjective groups to represent for entire music clip.

Table I: The nine groups of Updated Hevner labels

Groups	**Adjectives in Groups**
Group 1	Bright, Cheerful, Happy, Joyous;
Group 2	Humorous, Light, Lyrical, Merry, Playful
Group 3	Calm, Delicate, Graceful, Quiet, Relaxed, Serene, Soothing, Tender, Tranquil;
Group 4	Dreamy, Sentimental
Group 5	Dark, Depressing, Gloomy, Melancholy, Mournful, Sad, Solemn;
Group 6	Heavy, Majestic, Sacred, Serious, Spiritual, Vigorous
Group 7	Tragic, Yearning;
Group 8	Agitated, Angry, Restless, Tense
Group 9	Dramatic, Exciting, Exhilarated, Passionate, Sensational, Soaring, Triumphant

Table II: Updated Hevner Model Vs 2 Dimension Quadrants

Group	2 D Quadrant
Happy	1
Merry	1
Calm	4
Dreamy	4
Sad	3
Sacred	4
Tragic	3
Angry	2
Exciting	1

Kaminskyj *et al*. [17] represents the correlation between different mood taxonomy models, Thayer's 2D emotion plane as well as updated Hevner adjective labels [10]. 46 adjectives fall into nine groups of the updated Hevner labels are shown in Table I. Table II shows the correlation of Thayer's 2D model and updated Hevner adjective model. The mood taxonomy model in this system combines these two models which are presented by [17] and this proposed model is shown in Fig. 1.

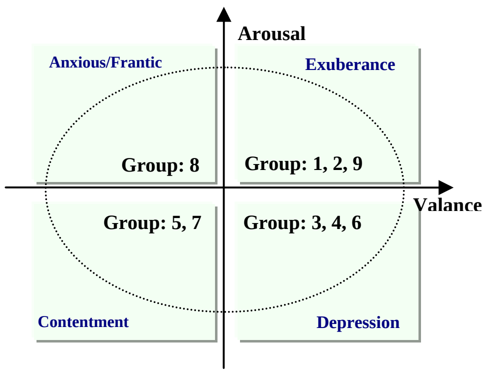

Fig.1. Proposed Mood Model

4. System Architecture

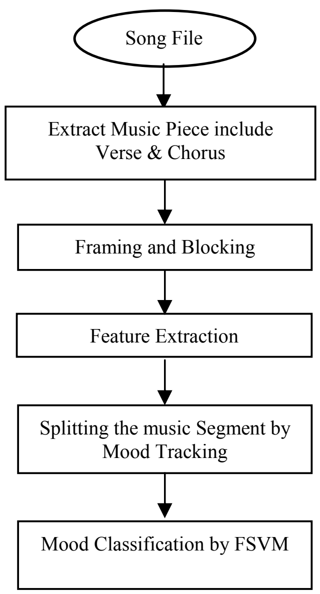

Fig.2. Overview of the proposed System

The proposed system first extracts music piece including verse and chorus parts manually. After framing and blocking, the feature sets are extracted from the entire music clip. By using intensity features, homogenous mood pattern is segmented from the music clip which we extract manually. This proposed system adopts image region growing methods for segmenting the music which contains the homogenous mood pattern. The feature sets of these separated parts are injected to the Fuzzy-SVM classifiers. Fig. 2 shows the overview of the proposed system.

4.1. Audio Data Set and Feature Extraction

Table III: The list of extracted feature sets

Features Name	
Intensity Features	Intensity
	Intensity Ratio
Timbre Features	Brightness
	Bandwidth
	Roll Off
	Spectral Flux
	Sub-band Peak
	Sub-band Valley
	Sub-band Contrast
Rhythmic Features	Rhythm Strength
	Average Correlation Peak
	Ratio between average peak and valley strength
	Average Tempo
	Average Onset Frequency

The data set used for this work consists of 60 songs proposed by Yang *et al*. [4] which are collected 60 famous popular songs from English Albums and annotated by 2Des model followed by updated Hevner adjective list. Each input song clip contains verse and chorus parts from the whole song, which is trimmed around one and a half minute. Before extracting feature, each music clip is first down sampled into uniform format: 16 kHz, 16 bits, mono channel and divided into non-overlapping frames of 32-ms length.

Mode, intensity, timbre and rhythm are of great significance in arousing different music moods in previous work. Juslin [14] also found that tempo, sound level, spectrum, and articulation are highly related to various emotional expressions. These findings are very similar although the exact words are different, such as rhythm versus tempo, and intensity versus sound level. This system extracted the recommended feature sets by Liu *et al*. [1]; intensity, timbre and rhythm. All the calculated feature sets are listed in Table III. Details on the computation of all the listed features can be found in [1].

4.2. Mood Tracking

In literature, the mood from some pieces of music may well change one or more times within a single piece. Therefore, it is not appropriate to detect an exclusive mood for an entire piece of music. Since the intensity, timbre, and rhythm are main primitives in mood detection, their changes also provide the main cues for a new mood event. Therefore, all of these three feature sets are used to complement each other in mood boundary detection [1]. In [1], an intensity outline is first implemented to coarsely detect potential boundaries, and then the timbre and rhythm features are used to detect possible mood changes in each contour of the intensity outline. These two steps are similar to the hierarchical process in mood detection discussed in the above section. It is noted that most of the boundaries detected by timbre and rhythm are matched with those from intensity outlines.

Therefore, in this mood tracking system rely on single feature known intensity. This system adopts image region growing method [18] for making split or merge decision upon the entire music clip.

4.3. Multi-Label Music Mood Classification

Based on our new mood taxonomy model, a hierarchical framework is proposed for mood classification, as illustrated in Fig. 3.

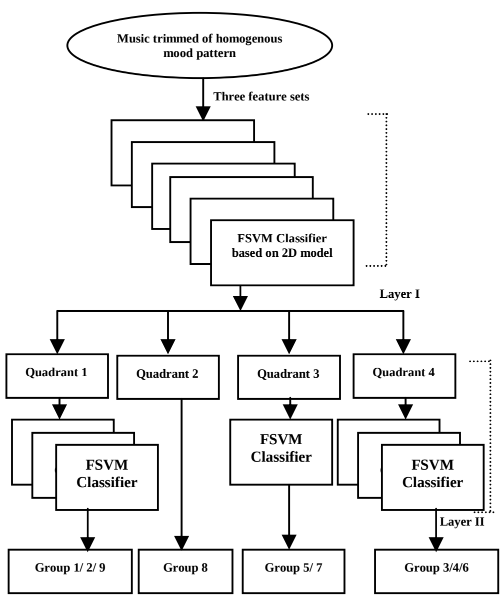

Fig.3. the proposed hierarchy framework

While support vector machines (SVMs) improved traditional perceptrons by using a higher dimensional space and identifying planes that provide maximal separation between two classes, they are essential binary classifiers. In order to increase the applicability of SVMs, it is necessary to extend them to multi-classification. Knerr, *et al*. [21] proposed O-A-O approach, whereby a separate SVM is constructed for every pair of classes. The O-A-O approach involves creating binary classifiers for each pair of classes, thus creating $N \times (N - 1)/2$ classifiers. Therefore, in this 9 class classification case needs 36 classifiers are needed. However, in our proposed hierarchy framework based on new mood taxonomy model needs only 13 classifiers for 9 class classification.

Layer I classification in Fig. 3, firstly classify the music trimmed into 4 quadrants based on Thayer's 2D model. The six FSVM classifiers are needed for four class classification. After classifying the entire music trimmed into four quadrants, layer II classification is followed. The requirement of classifier used is vary among each quadrant. After layer II classification, each of the music trimmed is annotated by the list of adjectives in updated Hevner model.

This approach uses fuzzy membership to each input point and reformulates SVM such that different input points can make different contributions to the learning of decision surface proposed by Lin *et al*. [12].

5. Conclusion

Although, the processing of one and half minutes music trimmed cost too much computational overhead and complexity, it may lead to more accurate result. This paper proposes the new mood taxonomy model and hierarchical mood classification framework to classify music mood. This framework can reduce computational complexity in considerable amount. Although, this system currently tests the 60 songs for multi-label music mood classification, the further classification with a large song dataset is required tend to be better accuracy. In conclusion, the proposed approach is a preliminary attempt with acceptable result for multi-label music mood classification from acoustic music data.

6. References

[1] L. Lu, D. Liu, H.-J. Zhang, "Automatic mood detection and tracking of music audio signals," IEEE Trans. Audio, Speech, and Language Processing, vol. 14*, no. 1*, pp. 5–18, 2006.

[2] Y. Feng, Y. Zhuang, Y. Pan, "Music information retrieval by detecting mood via computational media aesthetics," Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence Washington, DC, USA (2003).

[3] W. Wu and L. Xie, “Discriminating Mood Taxonomy of Chinese Traditional Music and Western Classical Music with Content Feature Sets,” Proceedings of the 2008 Congress on Image and Signal Processing, Vol. 5 - Pages 148-152, May 2008.

[4] Y.-H. Yang, Y.-F. Su, Y.-C. Lin, and H.-H. Chen, “Music emotion recognition: The role of individuality,” Proc. ACM SIGMM Int. Workshop on Human-centered Multimedia 2007, in [5]conjunction with ACM Multimedia (ACMMM/HCM'07), pp. 13–21.

[6] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H.-H. Chen, “A regression approach to music emotion recognition," IEEE Transactions on Audio, Speech and Language Processing (TASLP), 16(2):448–457, February 2008.

[7] A. Wieczorkowska, “Extracting Emotion From Music Data,” 2007.

[8] T. Li, M. Ogihara, “Content-based music similarity search and emotion detection,” Proc. ICASSP, pp. 17–21, 2006.

[9] K. Trohidis, G. Thoumakas, G. Kalliris, I. Vlahavas, “Multilabel Classification of Music into Emotions,” Proc. 9th International Conference on Music Information Retrieval (ISMIR) 2008.

[10] E. Schubert, “Update of the Hevner adjective checklist,” Perceptual and Motor Skills, 96, pp. 1117-1122, 2003.

[11] Knerr, S., Personnaz, L., Dreyfus, G., “Single-layer learning revisited: A step- wise procedure for building and training a neural network,” In Fogelman-Soulie and Herault, editors, Neurocomputing: Algorithms, Architectures and Applications, NATO ASI. Springer (1990).

[12] C. F. Lin, S. E. Wang, “Fuzzy support vector machines,” IEEE transactions on Neural Networks, 13(2), 464-471, 2002.

[13] C. L. Krumhansl, “*Music: A Link Between Cognition and Emotion*”. Current Directions in Psychological Science, 2002.

[14] P. N. Juslin, “Cue utilization in communication of emotion in musicperformance: relating performance to perception,” J. Exper. Psychol.:Human Percept. Perf., vol. 16, no. 6, pp. 1797–1813, 2000.

[15] PsySound, http://members.tripod.com/~densil/.

[16] Marsyas, http://marsyas.sness.net/.

[17] I. Kaminskyj, A. Uitdenbogerd, “A study of human mood tagging of musical pieces,” The inaugural International Conference on Music Communication Science 5-7 December 2007, Sydney, Australia.

[18] http://en.wikipedia.org/wiki/Region_growing.