An Approach for Mulit-Label Music Mood Classification

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

Moe Pwint
University of Computer Studies, Yangon
Myanmar
moepwint@gmail.com

***Abstract*—Music can express emotion in succinctly but in an effective way. Peoples select different music at different time concordance with listening time's mood and objectives. Music classification and retrieval by perceived emotion is natural and functionally powerful. Since, human perception of music mood varies individual to individual; multi-label music mood classification has become a challenging problem. Because music mood may well change one or more times in an entire music clip, an exact song may offer more than one music taste to the music listener. Therefore, tracking mood changes in an entire music clip is given precedence in multi-label music mood classification tasks. This paper presents self-colored music mood segmentation and a hierarchical framework based on new mood taxonomy model to automate the task of multi-label music mood classification. The proposed mood taxonomy model combines Thayer's 2 Dimension (2D) model and Schubert's Updated Hevner adjective Model (UHM) to mitigate the probability of error causing by classifying upon maximally 4 class classification from 9. The verse and chorus parts approximately 50 to 110 sec of the whole songs is exerted manually as input music trims in this system. Consecutive self-colored mood is segmented by the image region growing method. The extracted feature sets from these segmented music pieces are ready to inject the Fuzzy Support Vector Machine (FSVM) for classification. One-against-one (O-A-O) multi-class classification method are used, for 9 class classification upon updated Hevner labeling. The hierarchical framework with new mood taxonomy model has the advantage of reducing computational complexity due to the number of classifiers employed for O-A-O approach as only 19 instead of 36 classifiers.**

Keywords- self-colored music mood segmentation, music mood, music emotion

I. INTRODUCTION

Music is not merely a form of entertainment but also the easiest way of communication among people, a medium to share emotions and a place to keep emotions and memories. Booming of the Internet technology, there is more and more music on personal computer, in the music libraries and on the Internet. Therefore, automatic music analysis system such as music classification, music browsing and play list generation system are urgently required for music management facility. Because of various listening objectives in different time concordance, music classification and retrieval based on perceive emotion is mightily powerful than other tagging such as artist, album, tempo and genre. However, to the best of our knowledge, few systems claim to be able to automatically retrieval music by mood because of having two obstacles lies on this approach: one is that there is no perfect computational mood model yet, and the other is that mood is an item related with cultural background and involved scenario.

Though some psychological researches [1], [2] have been studied the relationship between music and perceived emotion for decades, the boom of music emotion classification can be dated back to within 10 years [3-8]. Nowadays, multi-label music mood classification has become very important due to the variety of individual preferences and the song being able to provide different mood swings. This paper is comprised of two parts; mood taxonomy model for decent mood annotation and music mood classification. A new mood taxonomy model by combining dimension model and adjective model is proposed for music mood annotation. Mood classification task is divided into two steps as self-colored music mood segmentation and mood classification upon the segmented music trimmed. Image region growing method is proposed for self-colored mood segmentation before classification. This may effectively assist for multi-label music mood classification. Then hierarchical classification framework has been proposed based on new taxonomy model to mitigate the number of classifier used from 36 to 19. Moreover, this may also reduce probability of error occurring by classifying upon 9 class classification to maximally 4.

II. RELATED WORK

It is usually argued that the mood in music is too subjective to be detected. However, as much research such as [1], [2] indicated that these emotions are able to be communicated. According to the literature, dimension base mood taxonomy model is more computation tractable than quite freely adjective model. However, dimension model cannot fulfill the users' distinctive nature and familiarity for annotation. Most of the previous multi-label classification task [6], [8] structured their system based on Farnsworth's adjective mood taxonomy model. This adjective model uses 23 adjectives categorized into 13 classes. Tao Li *et. al* [7] compared the classification accuracy between 13 classes and re-categorized them to 6 super-class. The 13 class classification accuracy is apparently lower than 6 class classification although using same feature sets and machine learning technique. Including the subjective nature of music and other difficulties of multi-class and multi-label

978-1-4244-6893-5/$26.00 © 2010 IEEE

classification derived very low accuracy in all of the multi-label music mood classification tasks.

Though the previous approaches [3-8] formed their system upon 20 to 30 sec music trimmed, and then they all agree that 20 to 30 sec trimmed has not sufficient to get the whole information from the song. Moreover, Lie Lu *et. al* [3] proved that mood in an entire music clip may well changeable. Therefore, mood tracking for entire song which containing the length more than 30 sec is crucial for multi-label music mood classification. This system proposed self-colored music mood segmentation via image region growing method followed by music mood classification. The region growing approaches [9-11], developed upon 2D or 3D images to enhance/filter of the desired images by seeded or unseeded region growing method. How they differ, the basic algorithm for their proposed method is same. They chose a point to grow known as seed point and finding the lowest difference among the neighbors from the seed point. Definite thresholds were assigned for deciding continuing grows or not.

Although mode, intensity, timbre and rhythm are of great significance in arousing different music mood, mode is very difficult to obtain from acoustic data. Therefore, only the features of intensity, timbre and rhythm are extracted which have been applied by Lie Lu *et. at* [3].

Various machine learning techniques has been proposed in previous work [3-8]. The scarcity of literature in multi-label classification, this system has been handled by decomposing the problem into a set of binary classification problems known O-A-O support vector machine (SVM). Moreover, this system intends to use fuzzy membership to each input point and reformulates SVM. The different input points can make different contributions to the learning of decision surface proposed by Chun-Fu Lin *et al*. [12].

III. Proposed System

In this proposed system, verse and chorus parts of the music piece is extracted manually. After framing and blocking, intensity features are extracted for homogenous mood segmenting from the original music trimmed.

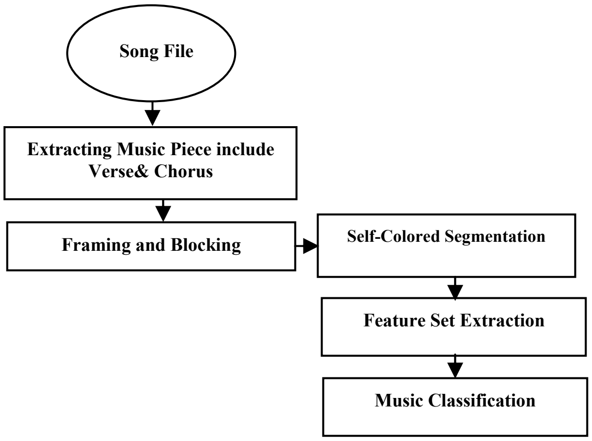


Figure 1. System Overview

This proposed system is adopted image region growing method for self-colored music mood segmenting. Then, the feature sets of these separated parts are injected into the FSVM classifiers. Figure 1 shows the overview of the proposed system.

IV. Dataset and Mood Taxonomy

Most of the single label classification such as [3-5] has been developed their model based upon Thayer's 2D model. The emotion classes are defined in terms of arousal/energetic (how exciting or calming) and valance/stress (how positive or negative). The four quadrants in Thayer's arousal and valance model are named exuberance, anxious/frantic, contentment and depression, respectively.

However, this dimension model confronts the lack of users' familiarity and verities of individual preferences, adjective models of mood taxonomy are adopted in most of the multi-label music mood framework. By far the most of popular checklist was that devised by Kate Hevner, formulated in 1930s, has been widely used to measure emotional responses to music. In 2003, Schubert [14] has been updated the Hevner adjective circle by refining her original words and addition some words to get the final list consisted of 46 words grouped into 9 clusters; Group 1: cheerful, bright, happy, joyous; Group 2: merry, humorous, light, lyrical, playful; Group 3: calm, delicate, graceful, quiet, relaxed, serene, soothing, tender, tranquil; Group 4: dreamy, sentimental; Group 5: sad, dark, depressing, gloomy, melancholy, mournful, solemn; Group 6: sacred, heavy, majestic, serious, spiritual, vigorous; Group 7: tragic, yearning; Group 8: agitated, angry, restless, tense; Group 9: exciting, dramatic, exhilarated, passionate, sensational, soaring, triumphant.

Ian Kaminskyj *et. al* [13] represented the correlation between different mood taxonomy models, Thayer's 2D emotion plane as well as Schubert's UHM [14]. The nine groups of updated Hevner's Adjective checklist are re-categorized via Thayer's 2D model.

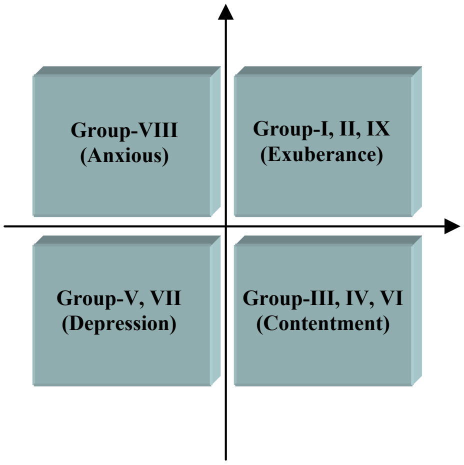

Group-VIII (Anxious)
Group-I, II, IX (Exuberance)
Group-V, VII (Depression)
Group-III, IV, VI (Contentment)

Figure 2. Proposed Framework

New mood taxonomy model is proposed in this system by combining dimension and adjective model according to their correlation as shown in Figure. 2.

A collection of 100 famous western popular songs are created as dataset of the proposed system. The various lengths from 50 up to 110 sec music trimmed are extracted manually.

Segmented self-colored music trimmed was labeled independently by a subject (30 year old female). Each music piece was labeled using Thayer's 2D model first and followed by UHM model annotating.

V. Audio Feature Extraction

Before extracting feature sets, 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. Intensity, timbre and rhythm feature sets developed by Lie Lu *et. al* [3] were extracted. Intensity is an essential feature in mood detection and it can be estimated using simple amplitude based measures. Spectral shape features and octave-based spectral contrast is calculated from the FFT. As rhythm is also closely related with people's mood response, rhythm strength, rhythm regularity and tempo are extracted from the segmented music piece. The detail calculation can be found in [3].

A summary of the used features are as follows. The two feature sets, intensity and intensity ratio are extracted to represent the songs intensity accounts. Then, bandwith, roll-off, spectral flux, sub-band peak, sub-band valley and sub-band contrast have been extracted to represent timbral difference among each song and the five feature sets of rhythm named rhythm strength, average co-relation peak, ratio between average peak and valley strength, average tempo and average onset frequency has also been extracted.

VI. Self-colored Music Mood Segmentation

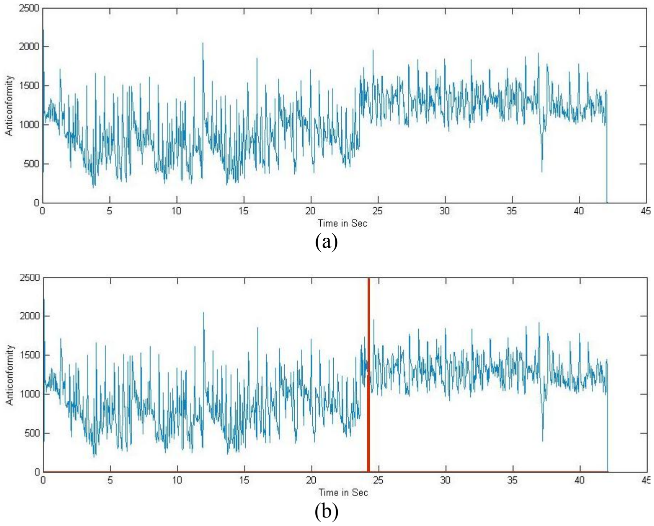
Music mood is a very subjective nature and entire song can bestow individual music tastes to individual listeners. Therefore, self-colored music mod segmentation plays the fundamental analysis for music mood classification. This proposed framework intended to segment between music mood changes in an entire music trimmed by using simple image region growing method. The four moods represented in Thayer's model are considered as a basic mood of the proposed system. Music mood segmentation decision is empirically considered upon two basic threshold values, Tr_1 and Tr_2 over song intensity feature. The absolute difference between mean of first 20% of the entire music clip and mean of last 20% of the same clip, multiplied by C_1% is defined as Tr_1. Mean of last 20% of the music clip is also multiplied by C_2% for Tr_2. Some enhancements are needed to segment upon 1 dimension intensity feature since almost all the region growing methods are applied for 2 or 3 dimension images [8-10]. Therefore, this system assigned seeded region instead of assigning a single seed point. Detail algorithm of self-colored music mood segmentation based on region growing method is shown in Figure 3.

```
Begin
    Get intensity from an entire music
    Assign Tr1;
    Assign Tr2;
    seedRegion = 0;
    N = length of the whole music clip in sec

    Repeat
        $Frame_i$ = Framing every 1 sec of music frame;
        Begin
            While mean($Frame_i$)<Tr1
                Grow;
                Increase i;
            End
            Reassign seedRegion;
        End
        If (length of growing frame > S sec) and seedRegion < Tr2
            Segmented;
        If (length of growing frame > S sec) and seedRegion > Tr2
            Segmented;
        Else
            Ignore;
        End
    Until (i != N)
End
```

Figure 3. Algorithm of Music Mood Tracking

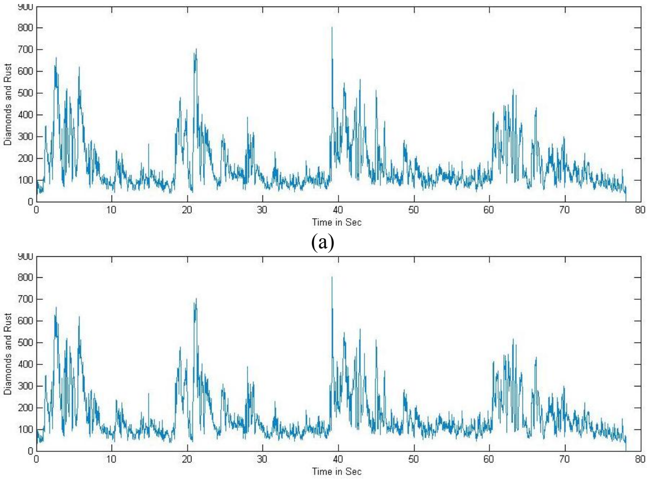
The experiments by testing with 110 songs (66 Western Popular songs and 44 Myanmar Popular Songs) can be proved that the proposed method has been working well. Some segmentation results for music mood classification are as follow. In figure 3, the changes mood in the entire song can be tracked and segmented well. Likewise, the arousal of constant mood with repeated pattern can be tracked well.



(a)

(b)

Figure 4. Music Mood Tracking and Segmentation (a) Original Intensity (b) After Tracking and Segmentation Process



(a)

(b)

Figure 5. Music Mood Tracking and Segmentation (a) Original Intensity (b) After Tracking and Segmentation Process

Decision of segmentation is based on four basic mood accounts in Thayer's 2D model. The presented mood change pattern shown in Figure 4 represents both depression and anxious mood patterns. Most the changes of low to high energy can be detected well by the proposed approach. It is found that most of over segmented pieces fall into exuberance quadrant. For contentment music trimmed as in Figure 5, this mood can track correctly. The performance of this approach can be improved, if there is a mechanism for detecting the changes for slowly varying intensity pattern with a fine tuning in setting thresholds.

VII. Multi-Label Music Mood Classification

All the segmented music pieces by self-colored music mood tracking are followed by annotation of music via human subjects, is based on proposed new mood taxonomy model. Then, features are extracted from each segmented music pieces. After this, each feature sets are ready to inject to the FSVM classifiers. It is necessary to extend SVMs to multi-classification while they are originally proposed for binary classification. One of the multi-class classification method O-A-O approach is applied in this framework. The O-A-O approach involves creating binary classifiers for each pair of classes, thus creating N × (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, only 19 classifiers are needed for 9 class classification.

We choose the appropriate fuzzy membership function generating method, better suited for desired application. Distance between points and class center base fuzzy membership function is calculated in this system as

$$s_i = 1 - \left| x_+ - x_i \right| / r_+ + \delta \qquad \text{If } y_i = 1 \qquad (1)$$

$$s_i = 1 - \left| x_- - x_i \right| / r_- + \delta \qquad \text{If } y_i = -1 \qquad (2)$$

where x_+ and x_- means of data points which $y_i = 1$ and $y_i = -1$ respectively. The radius of the class +1

$$r_+ = \max_{\{x_i | y_i = +1\}} \left| x_+ - x_i \right| \qquad (3)$$

and the radius of the class -1

$$r_- = \max_{\{x_i | y_i = -1\}} \left| x_- - x_i \right| \qquad (4)$$

respectively.

A hundred western popular songs, which are correctly segmented from Section VI, have been assumed the datasets of classification in layer I and 25 songs are employer in each quadrant. Accuracy after 5 fold cross validation in each class is shown in Figure 6.

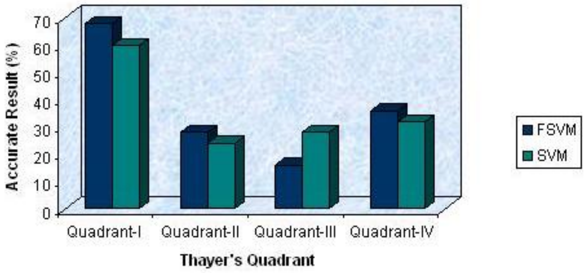


Figure 6. Performance Comparison between FSVM and SVM

Table I also lists the classification accuracy of proposed framework using fuzzy based SVM and SVM only.

TABLE I. Classification Accuracy of Proposed Framework

	Fyzzy SVM	**SVM**
Quadrant I	68	60
Quadrant II	28	24
Quadrant III	16	28
Quadrant IV	36	32

The performance of FSVM classifier is found to be 37% and SVM 36% respectively. According to this result, most of the songs in quadrant I is classified correctly and less accurate in quadrant III. In quadrant I and IV, the performance of FSVMs has higher accuracy than in ordinary SVM and only quadrant III in FSVM is less accuracy than in SVM.

VIII. Conclusion and Future Work

The output of the layer I classification will have been followed by layer II classification. Therefore, better accuracy is demanded in layer I classification to improve overall accuracy of the proposed system. Although a fully satisfactory result has not been acquired, FSVM has proved to provide superior accuracy over SVM. The result presented are only from preliminary experiments containing 20 songs for each group training data are not sufficient for this classification task and extended dataset is necessary for trusted classification result. According to this preliminary result, it is expected to obtain a better and robust classification through extended data sets.

REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] Y. H. Yang, 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.

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

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

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

[9] R. Pohle, K. D. Toennies, “Segmentation of Medical Images Using Adaptive Region Growing,” SPIE, 1337-46.

[10] Z. Lin, J. Jinn, H. Talbot, “Unseeded region growing for 3D image segmentation,” Proc., Selected papers from PanSydney Area Worksh. on Visual Information Processing (VIP2000), Sydney, Australia, 31–37.

[11] S. A. Hojjattolesami, J. Kittler, “Region growing – a new approach,” IEEE Trans. Image Processing, vol. 7, no. 7, pp. 1079–1084, July 1998.

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

[13] 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.

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