

Emotion-based Music Recommendation By Association Discovery from Film Music

Fang-Fei Kuo¹, Meng-Fen Chiang², Man-Kwan Shan² and Suh-Yin Lee¹

¹Department of Computer Science and Information Engineering, National Chiao-Tung University
Hsinchu, Taiwan

{ffkuo, sylee}@csie.nctu.edu.tw

²Department of Computer Science, National Cheng-Chi University
Taipei, Taiwan

{g9309, mkshan}@cs.nccu.edu.tw

ABSTRACT

With the growth of digital music, the development of music recommendation is helpful for users. The existing recommendation approaches are based on the users' preference on music. However, sometimes, recommending music according to the emotion is needed. In this paper, we propose a novel model for emotion-based music recommendation, which is based on the association discovery from film music. We investigated the music feature extraction and modified the affinity graph for association discovery between emotions and music features. Experimental result shows that the proposed approach achieves 85% accuracy in average.

Categories and Subject Descriptors

H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing – *methodologies and techniques*; J.4 [Computer Applications]: Social and Behavioral Sciences – *psychology*.

General Terms

Algorithms, Human Factors.

Keywords

music recommendation, emotion, affinity graph, association discovery

1. INTRODUCTION

With the development of digital music technology, it is essential to develop the music recommendation system which recommends music for users. Some work has been done on the personalized music recommendation to recommend based on the users' preference [1][2][11]. There exist two major approaches for the personalized music recommendation. One is the content-based filtering approach which analyzes the content of music that users liked in the past and recommends the music with relevant content. The other is the collaborative filtering approach which recommends music that peer group of similar preference liked.

Both recommendation approaches are based on the users'

preferences observed from the listening behavior. However, sometimes, it is more adequate to recommend music based on the emotions. Potential applications of emotion-based music recommendation includes of music score selection for production of home video, background music playing in shopping mall to stimulate sales, music playing in context-aware home to accommodate to inhabitant's emotion, and music therapy.

Most people experience music every day with affective response. For example, joy when listening to an excellent performance at a concert, sadness when listening to the music of a late night movie. Some researchers have devoted to understanding the relationships between music and emotion from the philosophical, musicological, psychological and anthropological perspectives [3]. To recommend music based on emotions, the straightforward approach is to recommend music by the rules, in terms of the relationship between emotion and music elements, observed by the psychological research. Another possible approach is to learn the rules by training from music labeled with emotion types. However, this approach is time-consuming.

In our work, to avoid labor work of emotion labeling, we propose a generic emotion-based music recommendation model to recommend music by association discovery from film music. In particular, we investigated music feature extraction and modify the *affinity graph* to discover the relationship between music features and emotions from film music. Experimental result shows that the proposed approach achieves 85% accuracy in average.

2. RELATED WORK

Ringo is a pioneering music recommendation system based on the collaborative filtering approach [11]. In Ringo, the preference of a user is acquired by the user's rating of music. Similar users are identified by comparing the preferences. Ringo predicts the preference of new music for a user by computing a weighted average of all ratings given by peer group of similar preference. MusicCat is a music recommendation agent based on user modeling [1]. In MusicCat, the user model is defined by the user.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'05, November 6–11, 2005, Singapore.

Copyright 2005 ACM 1-59593-044-2/05/0011...\$5.00.

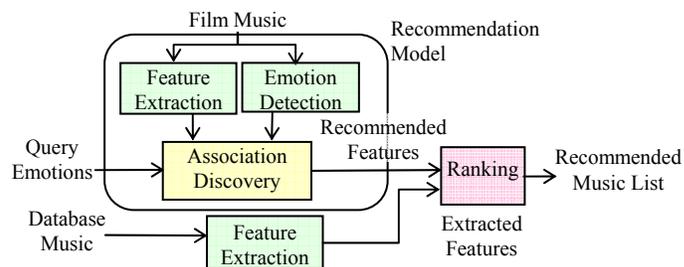


Figure 1. The proposed music recommendation process.

It contains information about user’s habit, preference or user-defined features, etc. MusicCat can automatically choose music from user’s music collection according to user’s model. MRS is a system which provides music recommendation based on music grouping and user interests [2]. It combines both content-based and collaborative filtering techniques to recommend music.

3. THE PROPOSED EMOTION-BASED MUSIC RECOMMENDATION MODEL

3.1 Model

Figure 1 shows the process of proposed generic music recommendation model. The heart of our approach is the construction of the recommendation model from film music. In films, music can serve as the overture to suggest whole film’s theme or spirit. The emotions, thoughts, wishes and characterizations of the characters can be expressed by music. Music can also change the audiences’ emotions and can be used as forewarning. Moreover, music can suggest situations, classes or ethnic groups. Music can neutralize or even reverse the predominant mood of a scene [4]. Kalinak has claimed that music is “the most efficient code” for emotional expression in film [5].

To construct the music recommendation model, we propose the process including the extraction of the music features from film music, the detection of emotions from film segment and the discovery of the associations between music features and emotions. Given the query emotions, the recommendation model will return the recommended music features with respect to the query. The recommended features are then employed to rank the database music and to recommend music for query emotions.

3.2 Emotion Detection

Many studies have been done on emotion detection from facial, voice, physiological signals and text. Especially, there exists research which extracts emotions from audio description in films provided for visually impaired people [10]. In this work, a list of emotion tokens was generated for each of the selected 22 types of emotions. Occurrences of emotion tokens are the indication of emotion being depicted. Recently, the problem of color-mood analysis of films based on syntactic and psychological models has been investigated [12]. All these studies may be utilized to help for detecting emotions from captions, visual features of scenes or acoustic features of dialogs in films. In this paper, we do not address the issues of emotion detection. We assume that the emotions associated with film music have been detected.

3.3 Music Feature Extraction

Music elements which affect the emotion include melody, rhythm, tempo, mode, key, harmony, dynamics and tone-color. Among these music elements, melody, mode, tempo and rhythm have stronger effects on emotions. Generally speaking, major is brighter, happier than minor; rapid tempo is more exciting or tenser than slow tempo.

Take Schubert’s *Der Lindenbaum* from *Winterreise* cycle as an example. First, Schubert used E major scale to express memory of the warm past (Fig. 2a.) Then, it is modulated from E major to minor to reflect the mournful situation of the wanderer (Fig. 2b.)

Sometimes emotion conveyed by music cannot be identified using only one of the above elements. For example, the music in minor



Figure 2. Excerpts of *Der Lindenbaum*, in Schubert’s *Winterreise*.

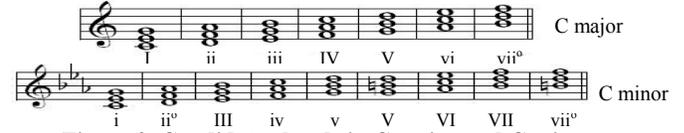


Figure 3. Candidate chords in C major and C minor.

scale but sprightly rhythm may be joyful rather than sad. Consequently, we consider the effect of the combination of three types of features and corresponding feature extraction algorithms related to these music elements.

Melody is the most important and memorable element in music. In our previous work on music style recommendation, we utilized chord as the melody feature for representing the music style and proposed method for assigning chords for melody [6]. Nevertheless, to aim at the music emotion, we modified the chord assignment algorithm to consider mode and key of the melody.

To assign the chord, the original polyphonic music should be pre-processed to obtain the main melody sequence and key/mode information. Some melody extraction algorithm can be used for MIDI files, such as all-mono [6]. Then, we determine the key signature of the music from MIDI key signature events.

Our modified chord assignment algorithm is a heuristic method based on the music theory and Harmony. The algorithm selects suitable chords from the candidates according to consonance and chord progression. The candidate chords we used here are the diatonic triads, which are basic and common chords. For minor scale, we selected 9 diatonic chords which are often used in composition from both natural and harmonic minor scales. Figure 3 shows two sets of the candidate chords in C major and C minor.

The algorithm includes two stages. In the first stage, melody is divided into parts first. Three rules are used to score candidates for each part. First, if the candidate has more notes which also appear in the part, it gets more points. Second, the longest note should be more dominant in the part; therefore, candidates that have the longest note get points. Last, tonic triad (I for major and i for minor) get more points in first and last parts, because music often begin and end at tonic triad. For each part, if the highest-score candidate is not unique, proceed to the second stage.

In the second stage, rules of chord progression which includes root motion and dissonance resolving are used. Root motion means the movement from one chord’s root to next chord’s root. We selected some common root motions for scoring, such as down a fifth (ex. I →IV) or up a second (IV→V). Then, some chords are unstable, and tend to resolve to more stable chords such as tonic triad. Therefore, if chord in previous part is unstable, some candidates that are more stable will get points. Finally, if the highest-score candidate is not unique, we assign a set of these candidates for each part, named as the chord-set.

Rhythm is the music feature that describes the timing information of music. Our rhythm extraction method includes the following steps: First, the beat sequence is extracted based on percussion instruments and is represented as a binary string, where one stands

for onset of the percussion note. For instance, a quarter note can be represented as 1000, where the basic unit is set to sixteenth note long. Then, the repeating patterns are discovered from the beat string using existing repeating pattern finding algorithm. The rhythmic pattern of music is the recurrent pattern with high frequency. We retained the highest frequency patterns for each music object.

In our approach, tempo is calculated from resolution of the music and beat density of the most repetitive pattern. The resolution of a music object is the number of ticks per beat. The following is the formula for calculating tempo:

$$\text{tempo} = \text{resolution} * NB / NS,$$

where NB is number of beat onset in a rhythmic pattern and NS is the length of the rhythmic pattern.

3.4 Association Discovery and Recommendation

Emotion-based music recommendation recommends the music corresponding to query emotions. More precisely, given a query set of emotions, we wish to find out the corresponding music features for recommendation. The relationship between music features and emotions of training data should be discovered. The graph-based approach, *Mixed Media Graph (MMG)* are adopted and modified for proposed emotion-based music recommendation.

3.4.1 Mixed Media Graph

MMG was proposed to find correlations across the media in a collection of multimedia objects [8]. A typical application of MMG is the automatic image captioning to automatically assign caption words to the query image. This is achieved by finding correlations between the image features and the caption words from a given collection of images and associated captions.

In MMG, all the objects and associated attributes are represented as vertices. For objects with n types of attributes, MMG will be an $(n+1)$ layered graph with n types of vertices and one more type of vertices for the objects. There are two types of edges in MMG. The *object-attribute-value link (OAV-link)* is the edge between an object vertex and an attribute vertex. The other type, *nearest neighbor link (NN-link)*, is the edge between two attribute vertices. An edge is constructed between each attribute vertex and each of its k nearest neighbors. After the construction of MMG, to find the correlations across the media, the mechanism of *random walk with restart* is employed to estimate the affinity of attribute vertices with respect to the query vertices. In detail, $f_q(v)$, the affinity of vertex v with respect to query vertex q is the steady-state probability that the random walker will reach v from q . In each vertex, the walker randomly selects and moves to the next vertex among the available edges with the exception to return to q with probability c . For the detail description of the computation of $f_q(v)$, refer to [8].

3.4.2 Music Affinity Graph

The music affinity graph is constructed as follows. For each trained music object, a music object vertex is created. For each music object vertex, four types of attribute vertices – emotion, chord, rhythm, and tempo vertices are created and attached. Note that one vertex is created for each of the perceived emotion while one vertex is created for each member of the set of chord-set

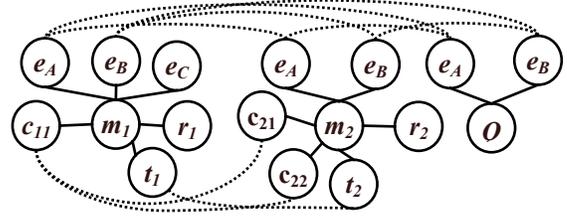


Figure 4. The music affinity graph G

extracted from a music object. The edge between chord (tempo) vertices is constructed based on the k -nearest neighboring while the edge between emotion (rhythm) vertices is constructed only when both vertices are of the same emotion (or rhythm).

Example Given a collection of two music objects $\{m_1, m_2\}$ in which music object m_1 , with chord feature $\{c_{11}\}$, tempo feature t_1 , rhythm feature r_1 , is perceived with emotions $\{e_A, e_B, e_C\}$ while music object m_2 , with chord feature $\{c_{21}, c_{22}\}$, tempo feature t_2 , rhythm feature r_2 , is perceived with emotions $\{e_A, e_B\}$. Figure 4 illustrates the constructed music affinity graph with respect to the query music Q and query emotions $\{e_A, e_B\}$ where the number of nearest-neighbors, k , is set to one.

The performance of the music affinity graph may be improved by the consideration of discrimination. The steady-state probability of an attribute vertex represents the affinity between the corresponding music feature value and the query emotions. However, it is not necessary to say that the music feature value with high affinity is highly correlative to the query emotions. It is possible that this music feature value also has high affinity with respect to other emotions. In other words, it is possible that there exists a feature value in spite of the emotion types. For example, a chord-set appears in both music with positive emotions and that with negative emotions. To address this problem, we propose the modified approach to accompany with *complement affinity graph G'* . The complement affinity graph G' is similar to the music affinity graph G except that all the nearest neighbor links from the query emotions are removed while the *complement query edges* are added. The complement query edge connects the query emotion vertex to the vertex, associated with music, of different emotion. After the construction of complement affinity graph, using the mechanism of random walk with restart, the affinity $g_q(v)$ for each vertex v in G' can be derived in the same manner of the affinity $f_q(v)$ in music affinity graph G . Consequently, the final affinity $h_q(v)$ is equal to $f_q(v) - g_q(v)$.

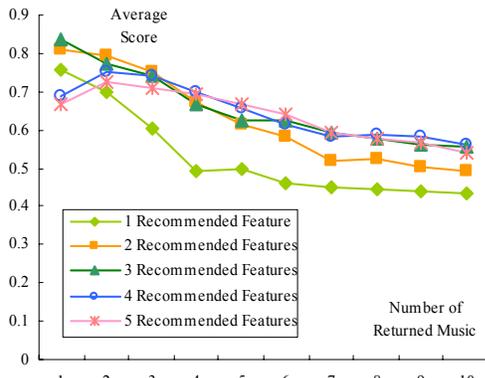
4. PERFORMANCE EVALUATION

To evaluate the effectiveness of our proposed music recommendation approach, we performed experiments on a collection of 107 film music from 20 animated films. We choose animated films because emotions in animated films are more clear and explicit in general. The 20 films include the productions of Disney, Studio Ghibli and DreamWorks, such as Lion King, Spirited Away, and Shrek. We collected the MIDI files of film music from the websites: <http://www.wingsee.com/ghibli/>, <http://www.ginevra2000.it/Disney/Midi/allmidi.htm>, and <http://www.hamienet.com>.

To simplify the experiments, the emotions of the music were annotated manually. The emotions used in our experiments were mainly selected from [9]. We added some emotions such as lonely

Table 1. Emotions used in the experiments

No.	Emotions
1	Hope, Joy, Happy, Gloating, Surprise, Excited
2	Love
3	Relief
4	Pride, Admiration
5	Gratitude
6	Gratification, Satisfaction
7	Distress, Sadness
8	Fear, Startle, Nervous
9	Pity
10	Resentment, Anger
11	Hate, Disgust
12	Disappointment, Remorse, Frustration
13	Shame, Reproach
14	Lonely
15	Anxious

**Figure 5. Performance of proposed approach (k=7, c=0.8)**

and nervous, and divided these emotions into 15 groups. These groups are shown in Table 1. Each MIDI file was annotated with one to seven emotion groups.

We take five-fold cross-validation in our experiments. In each test, the affinity graph was constructed from the training set and the emotions of one of the test set. In average, the music affinity graph contains 1000 chord-set nodes, 180 rhythm nodes, 86 tempo nodes and total 1600 nodes. All database music objects were ranked by the approach stated in section 3.4.2. Top-ten music were returned. The recommendation performance is measured by the similarity between query emotions and returned music's emotions. The performance measure used in our experiments is defined as

$$average_score = \frac{\sum_{i=1}^N Score_i}{N},$$

where N is the number of returned music, $Score_i$ is the similarity between emotion sets E_i of the i^{th} returned music and E_q of the query. $Score_i$ is defined as

$$Score_i = \frac{|E_i \cap E_q|}{\sqrt{|E_i| \times |E_q|}},$$

where $|E_i|$ is the cardinality of the set E_i , \cap is the set intersection operation. $Score_i = 1$ if E_i is the same as E_q .

Figure 5 shows the performance of the proposed recommendation approach (with $k = 7$ and $c = 0.8$.) The average scores of top-one music are above 0.8 using two or three recommended features. The overall average scores are above 0.5. The result using more than one recommended features are better; it is possible that

information of only one recommended feature is insufficient for recommendation.

5. CONCLUSIONS

In this paper, we presented a generic model to recommend music based on emotion. The core of our proposed approach is to construct the recommendation model from film music, for music plays an important role in conveying emotions in films. The model construction process consists of feature extraction, emotion detection and association discovery. We propose the feature extraction approaches to extract chord, rhythm and tempo, and modified the affinity graph approach to discover the associations between emotions and music features. Experimental result shows that the top-one result's average score achieves 85% using three recommended features.

6. REFERENCES

- [1] Chai, W. and Vercoe, B. Using User Models in Music Information Retrieval Systems. In *Proc. of Intl. Symposium on Music Information Retrieval (ISMIR'00)*, 2000.
- [2] Chen, H. C. and Chen, A. L. P. A Music Recommendation System Based on Music Data Grouping and User Interests. In *Proc. of ACM Intl. Conference on Information and Knowledge Management (CIKM'01)*, 2001.
- [3] Gabrielsson, A. and Lindstrom, E. The Influence of Musical Structure on Emotional Expression. In *Music and Emotion: Theory and Research*. Oxford University Press, 2001.
- [4] Giannetti, L. *Understanding Movies*. Prentice Hall, 2004.
- [5] Kalinak, K. *Settling the Score*. Madison, WI. University of Wisconsin Press, 1992.
- [6] Kuo, F. F. and Shan, M. K., A Personalized Music Filtering System Based on Melody Style Classification, In *Proc. of IEEE Intl. Conference on Data Mining (ICDM'02)*, 2002.
- [7] Ortony, A., Clore, G. L., and Collins, A. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.
- [8] Pan, J. Y., Yang, H. J., Faloutsos, C. and Duygulu, P. Automatic Multimedia Cross-modal Correlation Discovery. In *Proc. of ACM Intl. Conference on Knowledge Discovery and Data Mining (KDD'04)*, 2004.
- [9] Reilly, W. S. N. *Believable Social and Emotion Agents*. Ph.D. Dissertation, Carnegie Mellon University, 1996.
- [10] Salway, A., and Graham, M. Extracting Information about Emotions in Films. In *Proc. of ACM Multimedia Conference (MM'03)*, 2003.
- [11] Shardanand, U. and Maes, P. Social Information Filtering: Algorithms for Automating 'Word of Mouth', In *Proc. of the Conference on Human Factors in Computing Systems*, 1995.
- [12] Wei, C. Y., Dimitrova, N., and Chang, S. F. Color-Mood Analysis of Films Based on Syntactic and Psychological Models. In *Proc. of IEEE Intl. Conference on Multimedia and Expo (ICME'04)*, 2004.