Emotion-based Music Recommendation Using Audio Features and User Playlist

James J. DENG and Clement LEUNG
Department of Computer Science, Hong Kong Baptist University
Kowloon Tong, Hong Kong
{jdeng & clement}@comp.hkbu.edu.hk

***Abstract*—In this paper we utilize a dimensional emotion representation named Resonance-Arousal-Valence to express music emotion and inverse exponential function to represent emotion decay process. The relationship between acoustic features and their emotional impact reflection based on this representation has been well constructed. As music well expresses feelings, through the users' historical playlist in a session, we utilize the Conditional Random Fields to compute the probabilities of different emotion states, choosing the largest as the predicted user's emotion state. In order to recommend music based on the predicted user's emotion, we choose the optimized ranked music list that has the highest emotional similarities to the music invoking the predicted emotion state in the playlist for recommendation. We utilize our minimization iteration algorithm to assemble the optimized ranked recommended music list. The experiment results show that the proposed emotion-based music recommendation paradigm is effective to track the user's emotions and recommend music fitting his emotional state.**

***Index Terms*—Music emotion, music recommendation, conditional random fields, graph embedding, rank.**

I. Introduction

With the astounding growth of music on the Internet, music recommendation is important for discovering music that fit users taste or preferences. Current recommender systems are usually classified into three categories: content-based, collaborative filtering and hybrid approach [1]. For example, Pandora Radio utilizes content-based approach to recommend music with similar musical characteristics to the ones that the listener provided, while another popular music website Last.fm applies collaborative filtering approach to recommend music based on the listening behaviors of similar listeners. Though these approach suffer from respective problems, they all obtain good performances in their application environment.

As music is an art form and the soul of language which makes you feel a feeling or evokes your emotions, emotional expression in music is the key factor for music analysis. Thus emotion-based music recommendation provide a more natural and humanized way to help people better discovery music. However, We find that most of current music services ignore the emotion and sentiment influence or simply utilize tags to represent music emotion for example Steremood. With personal data becoming more and more important for personalized applications, users' history listening data such as music well expressing their feelings and providing a suitable data source to track their emotional states is important for emotion-based recommendation. This paper aims to provide emotion-based music recommendation that best fit listeners' emotions. As there exists some problems such as cold start, sparsity, and big data, we utilize the content-based approach to design emotion-based music recommendation.

In an attempt to well represent music emotion, we investigate related works on the area of perceived emotions on music, and utilize a dimensional emotion representation named Resonance-Arousal-Valence (RAV) to express music emotion. The relationship between acoustic features and their emotional impact based on this representation has been well established. As music well expresses feelings, through the users' historical playlist, we utilize the Conditional Random Fields (CRF) [7] to compute the probabilities of different emotion states, choosing the largest as the predicted user's emotion state. In order to recommend music based on the predicted user's emotion, we choose the optimized ranked list that has the highest emotional similarities to the music with the predicted emotion state. The minimization iteration method are utilized to return the optimized ranked recommended music list. The main contribution of this paper is that we propose to analysis the users' playlist in a session to implicitly track their emotions, and then recommend music that not only best fit their emotions, but also adoption of content-based approach.

The structure of this paper is as follows: in Section II we give a review of related works. In Section III we introduce our utilized music emotion model and describe emotion decay expression. Section IV explains emotion state prediction and presents our proposed emotion-based music recommendation approach. To this end, we describe our experiments and results in Section V, and finally give the conclusion in Section VI.

II. Related Work

In musicology and psychology, many researchers have already done a lot of works on recognition of emotion perceived from music. Currently there are two common theories to establish emotion models: discrete emotion theory and dimensional emotion theory [6]. The discrete emotion theory utilizes a number of emotional descriptors or adjectives to express basic emotions in human beings such as joy, sadness, and calm. The most widely used OCC [11] model hierarchically describes 22 emotion type specifications . The research community of Music Information Retrieval Evaluation eXchange (MIREX) also has classified music emotion into five categories by clustering different emotion labels [4]. However, the disadvantage of discrete emotion theory is that emotion terms or descriptors are ambiguous and not able to accurately describe emotions and measure their intensities. Conversely, the dimensional emotion theory believes that emotion should be regarded as continuous in a dimensional space. Thayer suggested that the two underlying dimensions were energetic arousal and stress [19], which is named arousal-valence emotion model widely utilized in music emotion recognition and proved effective in [9], [21]. Yang et al. [21] utilize the arousal-valence model to regress the music emotion as a point in the Cartesian space, and rank the music emotion for retrieval. The disadvantage of this two-dimensional model is not able to distinguish certain emotions. Further, a three-dimensional emotion model decomposes arousal into energy arousal and tension arousal in [13], which obtains the good performance in classification and regression. A layered model of affect [3] suggests map emotions represented by OCC model into a three dimensional PAD space, whose dimensions are pleasure, arousal and dominance, where the dominance represents the controlling and dominant nature of the emotion. However, the dominance is not suitable for representing one aspect of emotions invoked by music. In the domain of emotions induced by music, apart from pleasure and arousal, Bigand et al.[2] find that the third dimension seems to have an emotional character which measured by musicological features, like continuity-discontinuity or melodic-harmonic contrast.

Most of research work on emotion prediction are related to regression and Markov model. However, the limitation is that Markov model inferences the current emotion state depending on the previous emotion state. Lafferty et al. [7] proposed Conditional Random Field (CRF) to build probabilistic model, in which we can inference the listener's current emotion state depending not only on the previous series of states but also the following emotion states. A dynamic continuous factor graph model are utilized to predict user's emotion at specific time based on his historical information in [23]. Through the comparison of other prediction models, conditional random field approach has better performance in preliminary prediction. [14] utilizes CRF to model the relationship between the acoustic data and emotion parameters over time, which inspires our work for tracking listeners' emotion dynamically.

Though there are many works on music recommendation such as [16], [10], only a little papers are related with dynamic emotion-based music recommendation. [21] constructed a ranking-based emotion recognition system, which only used two dimensional valence and arousal cartesian space to represent emotion and obtained the relative emotional similar music. Furthermore, [8] proposed a personalized music recommendation combines the content-based, collaboration-based and emotion-based methods by computing the weights of the methods according to the listener's interests. Jun et al. [5] utilize Smith-Waterman (SW) algorithm to measure the similarity between the mood sequence and then recommend music list based on this similarity. However, these approaches still have problems to track listeners' emotion state and dynamically provide music recommendation fitting their emotions.

III. MUSIC EMOTION MODEL

As the two-dimensional Thayer's emotion model has already been proved effective in other researchers' work, in this paper we inherit the merits of this model. However, in the context of music psychology and philosophy, we find that emotional resonance is also an important measurement for deep emotion expression, and emotion induced by music seems to have an association with musicology, like continuity-discontinuity or melody-harmony [2], which evoke resonant or dissonant emotional response. Therefore, we give a new terminology "Resonance" for representing the third dimension. Therefore, the three-dimensional emotion model Resonance-Arousal-Valence (RAV) is utilized to represent music emotion, where arousal refers to whether the music activate your emotion or not, valence represents whether the music evoke your pleasure or displeasure emotions, and resonance stands for emotional resonance or emotional dissonance.

A. Music Emotion Representation

The Resonance-Arousal-Valence music emotion model can be regarded as a three-dimensional music emotion space, which combines the advantages of discrete and continuous representation of emotion. The emotion of a piece of music can be represented by $E = < s, rav >$, where s represents the emotion classification, and rav represents music emotion value, denoted by $rav = (\gamma, \alpha, \nu)$, which consists of resonance-based feature vector γ, arousal-based feature vector α and valence-based feature vector ν. The Section V-B describes these feature vectors in detail.

Suppose there are N pieces of music, we formulate a music-to-audio feature matrix $X_{L\times N} = [\mathbf{R} \mid \mathbf{A} \mid \mathbf{V}]^{\mathrm{T}}$. We aim to find the optimal matrix $A_{L\times d}$ to project the audio features to $\mathbb{R}^d$ by $X^{\mathrm{T}}A$. Graph Embedding attempts to find the projective map that optimal preserve the neighborhood structure of the original data set [20]. Given a graph $G = \{X, W\}$ with N vertices, each vertex represents a piece of music. Let W be a symmetric adjacency matrix with $W_{i,j}$ representing the weight, thus this graph embedding best preserves the relationship between different music in the dataset, we aim to find the optimal low dimensional representation for this graph.

Minimizing $\sum_{i,j} \|y_i - y_j\|^2 W_{i,j}$ can be transformed to the following optimization problem.

$$\begin{aligned} \hat{A} &= argmin\{Tr(A^{\mathrm{T}}XLX^{\mathrm{T}}A)\} \\ s.t. &\quad A^{\mathrm{T}}XDX^{\mathrm{T}}A = I \end{aligned}$$

where D is a diagonal matrix with $D_{i,i} = \Sigma_j W_{i,j}$, and $L = D - W$ is the graph Laplacian matrix, and I is the identity matrix. Therefore, after the computation of eigenvalues by $XLX^{\mathrm{T}}A = \Lambda XDX^{\mathrm{T}}A$, the optimal $\hat{A}$ are the eigenvectors corresponding to the maximum eigenvalue. As the similarity matrix W can be formulated by different similarity criteria such as Euclidean distance or cosine similarity *etc.*, currently there exist some popular dimension reduction methods such as Laplacian Eigenmap (LE), Locally Preserving Projections (LPP). We adjust similarity computation in music context for

TABLE I
EMOTION STATE (ES) OF THE RAV SPACE

ES	Octant	ES	Octant
S_1	+R+A+V	S_5	-R+A+V
S_2	+R+A-V	S_6	-R+A-V
S_3	+R-A-V	S_7	-R-A-V
S_4	+R-A+V	S_8	-R-A+V

$X^{\mathrm{T}}X$ to measure all pairwise music emotional similarities, and set nearest neighbors threshold to obtain W.

The simplest approach to obtain rav is to regress resonance-based features, arousal-based features and valence-based features, respectively, and confirm all their ranges from -1 to 1. Therefore, each piece of music is located in the point denoted by P of RAV space, then we define the initial intensity of emotion invoked by music based on its distance $|OP|$ from zero point denoted by $O(0,0,0)$ to point P. Suppose $|OP| = \rho$, thus the maximum value of $|OP|$ equals to $\sqrt{3}$. In order to simply describe emotion intensity without scale values, we divided the maximum emotion intensity into three extents to express different strengths such as slight, moderate and strong. An emotion state is modeled as a discrete state which stands for an octant in RAV space. As there are eight octants in the space, we suppose eight discrete emotion states, denoted by a collection S = $\{s_1, s_2, \cdots, s_8\}$. Table I shows the emotion states corresponding to the octants in RAV space.

As the intensity of emotion dynamically changes during its life cycle, the intensity of an emotion will to weaken through time after it is generated. In this paper, we follow the Picard's work in [12] to represent emotion intensity decay by an inverse exponential function. The relationship between the intensity of emotion denoted by e and the duration time denoted by t is given by

$$I(e,t) = I(e,t_0) \times \exp(-\omega \cdot t) \tag{1}$$

where constant ω represents decay factor of the emotion e, determining the speed of emotion decay. We utilize inverse sigmoid function to simulate emotion intensity decay [18].

$$I(q_i, t, t_0, \mu, \omega) = \frac{q_i}{1 + \exp(t - t_0 - \mu)\omega} \tag{2}$$

where q_i refers to the initial emotion intensity triggered at time t_0 which is the onset time of the emotion generated, ω is the decay factor of the emotion, and μ is the half-life time of the emotion. If the ω is smaller, the emotion intensity will decay more slowly, otherwise the emotion intensity will decay more fast. We found that negative emotions are more heavily weighted and lasting longer than positive emotions.

IV. EMOTION-BASED MUSIC RECOMMENDATION

We propose emotion-based music recommendation relaying on the session of the listener's historical playlist. Through the playlist of the listener u in a session, we implicitly obtain the corresponding emotion state sequence denoted by $S_u = \{s_u^{t_1}, s_u^{t_2}, \cdots, s_u^{t_i}\}$, which has great effects on the next emotion state $s_u^{t_{i+1}}$. Using the the expression of emotion

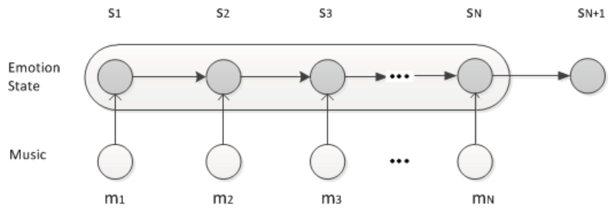

Fig. 1. Graphical model of a linear-chain CRF for music emotion state

intensity decay, we are able to compute emotion influence of each piece of music in the sequence to current emotion state respectively. Here, we utilize discriminative and conditional probability theory to obtain the output of the probability of current predicted emotion state, given the observations S_u. Accordingly, we define the emotion state influence weight proportion $P(s_u^{t_{i+j}}|s_u^{t_i})$, representing emotion state $s_u^{t_i}$ at time step t_i influencing the emotion state at the time step t_{i+j}. while the sum of influence weight proportions equals to one.

A. Emotion State Prediction

In order to recommend music based on the listener's time-varying emotion, we utilize Conditional Random Fields (CRF) to predict the listener's emotion state. CRF are undirected graphical and discriminative models, which represent the conditional probability $P(Y|X)$ of a particular label sequence Y given observation sequence X. Thus, the task of emotion state prediction can be expressed as $\hat{s} = argmax\ P(s_i|S_u)$, where $S_u = \{s_u^{t_1}, s_u^{t_2}, \cdots, s_u^{t_i}\}$ is a sequence of emotion states of user u, and s_i is a candidate emotion state, which belongs to the finite emotion states S = $\{s_1, s_2, \cdots, s_8\}$. After maximization of $P(s_i|S_u)$, we are able to obtain the best predicted emotion state $\hat{s}$ as the listener's current emotion state.

With the definition of CRF and in the scenario of the listener's music emotion state prediction, the formula of applied CRF to emotion state prediction is represented by

$$P(s_i|S_u) = \frac{1}{Z(S_u)} \exp\{\sum_{k=1}^{K} \lambda_k F_k(s_i, S_u)\} \tag{3}$$

where F_k represents feature function with λ_k whose weight is estimated from training data of emotion state transition probability, and K is the number of categories of emotion states in the listener's listening history record with $1 \leq K \leq 8$, and $Z(S_u)$ is a normalization factor that guarantees the distribution sum equal to one. The expression of $Z(S_u)$ is given by

$$Z(S_u) = \sum_{S} \exp\{\sum_{k=1}^{K} \lambda_k F_k(s_i, S_u)\} \tag{4}$$

Since previous emotion states have influence on the current emotion state, we use emotion intensity decay function to represent feature function in CRF. Therefore, the aggregated features $F_k(s_i, S_u)$ can be formulated by

$$F_k(s_i, S_u) = \sum_{j=1}^{i-1} \{f_k(t_i)\} = \sum_{j=1}^{i-1} \frac{q_i}{1 + exp(t_i - t_0 - \mu)\omega_k} \tag{5}$$

where $f_k(t)$ is the k-th emotion intensity decay function with the decay factor ω_k. Therefore, through above conditional random fields approach, initially, the listener's current emotion state $\hat{s}$ can be predicted.

B. Emotion-based Recommendation

Our music recommendation scheme is based on the predicted listener's current emotion state. Through the implicit sequence of emotion states, apart from we are able to predict the current listener's emotion state, we can also divide this sequence into subsequences based on the same emotion states. Given a user's playlist in a session denoted by $M_u = \{m_1, m_2, \cdots, m_N\}$, we implicitly obtain the corresponding emotion state sequence $S_u = \{s_1, s_2, \cdots, s_N\}$. Suppose the predicted emotion state is s_i, while there are k pieces of music belonging to this emotion state, thus we choose this subsequence denoted by S_{u,s_i} as the recommendation benchmark. Therefore, the proposed recommendation system attempts to make a decision that selects an optimized ranked music list that has the highest emotion similarities to the S_{u,s_i}. Given a piece of music m_i with corresponding audio features denoted by x_i, thus the emotion of m_i is represented by $x_i^{\mathrm{T}} A$. Thus we are able to compute the music emotion similarity by

$$\mathcal{F}_i(m_i) = (x_i^{\mathrm{T}} A)(x_j^{\mathrm{T}} A)^{\mathrm{T}} = x_i^{\mathrm{T}} A A^{\mathrm{T}} x_j \tag{6}$$

As each piece of music in the sequence S_{u,s_i} have different influence to current emotion state, the influence weight λ_{m_i} for music m_i is given by $\frac{I(t_{m_i})}{\sum_{j=1}^{k} I(t_{m_j})}$, where t_{m_i} refers to time step when the listener perceives emotion invoked by music m_i.

The ultimate goal to emotion-based recommendation is to return an optimized ranked list that has the highest emotional similarities to S_{u,s_i}. Therefore, we convert this rank problem to the optimization problem by minimizing emotion similarity between S_{u,s_i} and $\mathbb{N}$ candidate pieces of music.

$$\begin{aligned} &\min \sum_{j=1}^{\mathbb{N}} \Lambda \mathcal{F} + \frac{\eta}{2} \|\Theta\|_2^2 \\ &s.t. \quad \Theta = [\theta_1, \theta_2, \cdots, \theta_{\mathbb{N}}]^{\mathrm{T}} \end{aligned}$$

in which $\Lambda = [\lambda_1, \lambda_2, \cdots, \lambda_k]$, $\mathcal{F} = [\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_k]^{\mathrm{T}}$, and $\theta_i = m_i$. We utilize minimization iteration methods to solve the above rank problem. The Algorithm 1 describes ranking procedure for emotion-based music recommendation in detail.

Algorithm 1: Emotion-based Music Recommendation

Input: Sequence S_{u,s_i}, projection matrix A, regularization parameter η

Output: A ranked music list Θ_M with highest emotional similarities for recommendation

$\lambda_i \leftarrow \frac{I(t_{m_i})}{\sum_{j=1}^{k} I(t_{m_j})}$
$\mathcal{F}_i \leftarrow x_i^{\mathrm{T}} A A^{\mathrm{T}} x_j$
initialize $\Theta_M \leftarrow$ RandomCandidateSet($\mathbb{N}$)
$\Theta_M \leftarrow sort(\Theta_M)$
$\Phi(\Theta_M) \leftarrow \sum_{j=1}^{\mathbb{N}} \Lambda \mathcal{F} + \frac{\eta}{2} \|\Theta\|_2^2$
for $i = 1 \rightarrow N$ **do**
  if $\Phi(\Theta_{M_i}) < \Phi(\Theta_M)$ **then**
    replace $\theta_{M_i} \leftarrow \theta_i$
    update $\Theta_M \leftarrow sort(\Theta_{M_i})$
  end if
end for
return Θ_M

V. Experiment

A. Data Sets and Evaluation

In the experiment, we have carefully collected 275 western classical music clips from Amazon website free preview music and StockMusic website. The styles of these music clips contains concerto, sonata, symphony, and string quartet, and the total of 10 composers of these classical music are Bach, Beethoven, Brahms, Chopin, Haydn, Mozart, Schubert, Schumann, Tchaikovsky, and Vivaldi. Then we choose representative music excerpts which are able to invoke emotions as experimental data. All these digital files are converted to a uniform format, with sampling rate 22050 Hz, 16 bits, 705 bit rate, and stereo channel. Each experimental music excerpt is truncated to keep 30 seconds length.

We use Mean Average Precision (MAP) to measure the performance of emotion-based music ranking for recommendation. The mean average precision is based on the precision at position at k and average precision (AP). If there are r_k music emotional relevance in the top k recommended music list, then $P@k = \frac{r_k}{k}$, thus the AP is given by

$$AP(q) = \frac{1}{r_q} \sum_{i=1}^{k} \{P@k \times rel(k)\} \tag{7}$$

where $rel(k)$ is an indicator function equaling 1 if music at rank k is emotional relevant, otherwise zero. Thus MAP is obtained by meaning the AP.

B. Feature Extraction

In the related works on feature extraction, [9], [17], [22], [6], [15], we find that different music attributes such as timbre, rhythm, harmony and different acoustic characteristics such as energy, spectrum and tempo reflect different emotional expressions. The following parts give details about the arousal-based, valence-based and resonance-based features. Each 30 long seconds music excerpt is performed segmentation and frame decomposition with half overlapping. As for each frame, we extract arousal-based features, valence-based features, and resonance-based features, respectively. Finally we compute the statistical values of these selected features such as mean, standard variance and the difference. There are total 64 arousal-based features, 88 valence-based features, and 49 resonance-based features to extract in the experiment.

Arousal-based Features: Intensity or dynamic represents loudness, which is correlated to arousal such as high intensity arousing excited or joyful feelings or emotions, while low dynamics arousing neutral or depress emotions. The acoustic features often utilized to describe them is energy. The average energy of the given music is computed by root-mean-square (rms) method. Low energy and high energy are commonly used to express percentage of frames contrasted to average energy. High frequency energy measures the amount of energy above the cut-off certain frequency, which reflects the extent of brightness. In addition, since entropy is a useful tool to measure information, relative entropy of spectrum are also used to measure the degree of emotion arouse. Pitch represents fundamental frequency of a sound, thus high or low pitch represents different emotional expression such as active or inactive. In addition, chroma features are often utilized to describe energy distribution by twelve simitone (from A to G#) in Western music.

Valence-based features: Timbre is a key and comprehensive factor to express different emotional expressions. A special timbre maybe inspire valence response or unpleasant feelings from the listener. The acoustic features often utilized to represent timbre are Mel-frequency cepstral coefficients (MFCCs), and statistical spectrum descriptors (spectral shape and spectral contrast). Spectral shape features are usually obtained by short-overlapping frames through Hanning window and discrete Fourier transform (DFT). Spectral shape features are consist of spectral centroid, flux, flatness and rolloff which represents the frequency less than a specific proportion of spectral distribution. Spectral contrast features describe the comparison or correlation of spectrum, such as spectral kurtosis, valley, skewness, regularity, spread and zero-crossing rate. We represent timbre features by using above MFCCs and spectrum characteristics. Rhythm also has important effect on invoking pleasure or displeasure emotional feelings. For example, a particular rhythmic music may inspires valence or pleasure feelings, while non-rhythmic music often appears boredom or not feeling good. As rhythm reflects different duration over a steady background of the beat, which is related with characteristics such as beat spectrum, beat onsets, onset rate, silence rate, fluctuation, event density and tempo having contribution to express rhythm features.

Resonance-based features: Resonance means whether the music own melodic or harmonic properties invoking your emotional resonance or dissonance. In Western music, melodic intervals usually refers to separately played music notes, thus melodic intervals are able to described by ascending or descending musical intervals. Harmony refers to simultaneously performed tones or chords that represent mixture sounds such as muddy, sharp, and smooth. A harmony chord often reflects consonance, while an unharmony chord often reflects dissonance, thus they have effect on invoking your resonant or dissonant emotional feelings. Consonance features are often defined by the peaks of spectrum and their space of these spectral peaks, thus we use spectral peaks and roughness to describe music consonance characteristic. Tonality describes hierarchical pitch relationship between center key, thus key, key clarity and tonal fusion represented by frequency ration of the component tone, are often utilized to represent tonality. Based on the experience, we find that major mode often indicates emotional resonance, while minor mode often indicate disgust or dissonance emotional feelings. Therefore, melodic intervals and harmony related acoustic characteristics are utilized to represent resonance-based features in the paper.

TABLE II
EMOTION RECOGNITION OF PCA, SVR, AND GRAPH EMBEDDING

Method	75 Training	150 Training
PCA	77.3%	80.7%
SVR	65.3%	71.3%
$GraphEmbedding$	81.3%	85.3%

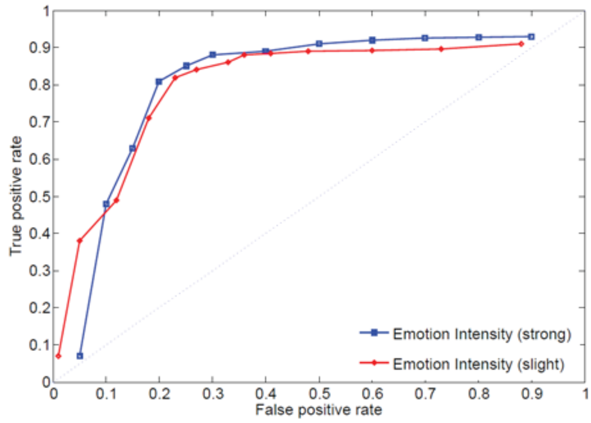

Fig. 2. ROC curve for emotion prediction accuracy with strong and slight emotion intensity samples

C. Results

We first evaluate the performance of music emotion representation. The experiments are conducted on the subdataset with 75 training dataset and 150 training dataset. Table II shows the results of PCA, SVR, and Graph Embedding for emotion recognition. It shows that graph embedding performs better than PCA and SVR in our dataset.

For listeners' emotion state prediction, we utilize CFR chain in Matlab to obtain the emotion state with the maximum possibility, then evaluate the predicted emotion state with the listener's real emotion state. Through multiple test of prediction on music with different emotion state, thus we can obtain the emotion state prediction accuracy. Figure 2 shows Receiver Operating Characteristic (ROC) curve for emotion state prediction accuracy with strong and slight emotion intensity samples, given the different decay factors ω corresponding to different emotion states. In the experiment, we use Maximum Likelihood Parameter (MLP) approach to estimate decay factors.

To evaluate the recommendation performance, we set $\eta = 0.01$ and $\mathbb{N} = 5$ to iteratively obtain the minimization of $\Phi(\Theta_M)$. The Table III shows the average precision for recommended music list. From the observations, it shows that the recommended music lists have obtained the satisfied results.

TABLE III
AVERAGE RANKING ACCURACY BY P @ K

P @ k	Average Precision
P@1	0.8750
P@2	0.8125
P@3	0.7500
P@4	0.7186
P@5	0.6500
MAP	0.7612

VI. CONCLUSION

In this paper, we utilize a dimensional emotion representation named Resonance-Arousal-Valence to express music emotion and inverse exponential function to represent emotion decay. The relationship between acoustic features and their emotional impact based on this model has been well constructed. As music well expresses feelings, through the users' historical playlist in the session, we utilize the Conditional Random Fields to compute the probabilities of different emotion states, choosing the largest as the predicted user's emotion state. In order to recommend music based on the predicted user's emotion, we choose optimized ranked list that has the highest emotional similarities to the music with the predicted emotion state for recommendation. The minimization iteration method are utilized to compute the optimized ranked recommended music list. The result show that the proposed approach is effective for personalized emotion-based music recommendation.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749, 2005.

[2] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19(8):1113–1139, 2005.

[3] P. Gebhard. Alma: a layered model of affect. In *Autonomous Agents & Multiagent Systems/Agent Theories, Architectures, and Languages*, pages 29–36, 2005.

[4] X. Hu and J. S. Downie. Exploring mood metadata: Relationships with genre, artist and usage metadata. In *International Symposium/Conference on Music Information Retrieval*.

[5] S. Jun, S. Rho, and E. Hwang. Music retrieval and recommendation scheme based on varying mood sequences. *Int. J. Semantic Web Inf. Syst.*, 6(2):1–16, 2010.

[6] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull. Music emotion recognition: a state of the art review. *Information Retrieval*, (Ismir):255–266, 2010.

[7] J. Lafferty, A. McCallum, and F. Pereira. *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, volume pages, pages 282–289. Citeseer, 2001.

[8] C. C. Lu and V. S. Tseng. A novel method for personalized music recommendation. *Expert Systems with Applications*, 36(6):10035–10044, 2009.

[9] L. L. L. Lu, D. Liu, and H.-J. Z. H.-J. Zhang. Automatic mood detection and tracking of music audio signals, 2006.

[10] B. McFee, L. Barrington, and G. R. G. Lanckriet. Learning content similarity for music recommendation. *IEEE Transactions on Audio, Speech & Language Processing*, 20(8):2207–2218, 2012.

[11] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*, volume 18. Cambridge University Press, 1988.

[12] R. W. Picard. *Affective Computing*, volume 136. MIT Press, 1997.

[13] U. Schimmack and R. Reisenzein. Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion Washington Dc*, 2(4):412–417, 2002.

[14] E. Schmidt and Y. Kim. Modeling musical emotion dynamics with conditional random fields. In *ISMIR'11*, pages 777–782, 2011.

[15] E. M. Schmidt, D. Turnbull, and Y. E. Kim. Feature selection for content-based , time-varying musical emotion regression categories and subject descriptors. *Spectrum*, pages 267–273, 2010.

[16] B. Shao, M. Ogihara, D. Wang, and T. Li. Music recommendation based on acoustic features and user access patterns. *IEEE Transactions on Audio, Speech & Language Processing*, 17(8):1602–1611, 2009.

[17] J. Shen, J. Shepherd, and A. H. H. Ngu. Towards effective content-based music retrieval with multiple acoustic feature combination, 2006.

[18] B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. A formal model of emotions : Integrating qualitative and quantitative aspects. *Language*, (September):1–5, 2008.

[19] R. Thayer. *The biopsychology of mood and arousal*. Oxford University Press, USA, 1989.

[20] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[21] Y. Yang and H. Chen. Ranking-based emotion recognition for music organization and retrieval. *IEEE Transactions On Audio Speech And Language Processing*, 19(99):762–774, 2010.

[22] Y.-h. Yang, Y.-f. Su, Y.-c. Lin, and H. H. Chen. Music emotion recognition : The role of individuality. *Work*, pages 13–21, 2007.

[23] Y. Zhang, J. Tang, J. Sun, Y. Chen, and J. Rao. Moodcast: Emotion prediction via dynamic continuous factor graph model. *2010 IEEE International Conference on Data Mining*, pages 1193–1198, 2010.