An Emotion-oriented Music Recommendation Algorithm
Fusing Rating and Trust

Jiwei Qin[1,2,3] , **Qinghua Zheng**[1,2]

1. Ministry of Education Key Lab For Intelligent Networks and Network Security, Xi'an 710049
2. Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049
3 .Center for Educational Technology, Xinjiang University,Urumqi 830046
Email: jwqin.2008@gmail.com, qhzheng@mail.xjtu.edu.cn

Feng Tian[*1]**, Deli Zheng**[1,2]

1. Ministry of Education Key Lab For Intelligent Networks and Network Security, Xi'an 710049
2. Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049
E-mail: ftian@sei.xjtu.edu.cn, delizheng.2009@gmail.com

Abstract

With the overwhelming increase of music, it has become difficult to find music which suits the taste of a listener who is in a certain state of emotion. Focusing on the listener's emotional state, this paper presents an emotion-oriented music recommendation algorithm. First, the listener's similarity is calculated by the rating value and the trust value. More specifically, based on the number of ratings, two thresholds are set to extend the calculation strategy of listener similarity weight to selectively use the trust value and correlation to the rating value. Second, because the music and the listener's emotional state are different objects, there is no obvious way to match one with the other. The listener's preference for emotional connotation of music is introduced to bridge the gap between the music and the listener, and that solves the issue of how to match the listener's emotion with music. Lastly, considering the difference of listener's perception of musical content and the complexity of the listener's emotional response, we propose a comprehensive measure to evaluate the accuracy, the coverage and the listener's satisfaction degree with the recommendation. Experimental results show that the presented algorithm comparing the collaborative filtering and trust-based recommendation, results in a tiny loss of accuracy with the improvement of larger coverage, thus not only obtaining a perfect tradeoff between accuracy and coverage, but also increasing the degree of listener satisfaction.

Keywords: Similarity weight; Emotion; Trust; Music recommendation.

1. Introduction

Music as an emotion-loaded form is regarded as a common entrainment resource in everyday life as well as a part of the psychological function of the individual[21]. It loads emotional content to augment or reduce the listener's current emotion[12]. However, with the overwhelming number of available music, it has become difficult to find the right music which fits the taste of a listener who is in a certain emotion state.

With the development of computer algorithm, music recommender systems started to emerge in 1990's[1]. Such a system aims to recommend music to a listener based on the listener's profile, and the features of music that this listener has chosen in the past[14]. These music recommender systems are usually divided into two groups: collaborative filtering system and content-based filtering system. The former is also known as the "word-of-mouth" method[22]. It uses the listener's music ratings as feedback to predict the ratings of the music that has

---
*Corresponding Author: ftian@sei.xjtu.edu.cn.

not been heard by the active listener; Last.fm[3] and Ringo[8] are two typical recommendation systems based on this collaborative filtering method. Content-based filtering such as Pandora[3] depends on the music labels (title, genre, etc.) and physical features (for instance, the acoustic signal) to choose interesting music for a listener.

However, the features used in these two types of recommendation systems are too general. They involve little personalization, because they do not take the listener's contextual information in finding the music. Contextual music recommendation has emerged in recent years to meet the listener's demand for personalized music. It integrates the listener's contextual information, such as the listener's emotional state, location and weather conditions, into recommendation to provide music suitable for the listener in a certain context[2].

We focus on the listener's emotional state and seek music that fits the taste of a listener who is in a certain emotion state. There are new challenges to matching the emotional state of listeners with music. The lack of unified emotional categories of music causes inadequacy of reference data. In addition, it is unclear that the complexity of a listener's perception of music's emotional connotation gives rise to an emotional change. In reality, the evaluation of the emotion-oriented music recommendation obtains from the listener's judgment in real experience, then existing measures cannot perform music recommendation with full success[12].

This paper presents an emotion-oriented music recommendation algorithm. We use the trust value and the rating value to measure the listener similarity based on two thresholds defined by the number of co-ratings. The listener's preference for emotional connotation of music to be served bridges the gap between the music and the listener's emotional state to find the music suitable for the listener's emotion. Meanwhile a comprehensive measurement is performed to evaluate the accuracy, the coverage and the listener's degree of satisfaction with emotion-oriented music recommendation.

The rest of the paper is organized as follows. Section 2 lists the related theories and research in the psychology research, as well as the recommendation algorithms. Section 3 describes the details of the architecture of the emotion-oriented music recommendation algorithm. In Section 4, we conduct two groups of experiments to evaluate our algorithm. Section 5 presents our conclusions.

2. Related Works

In this section, we introduce the theories behind the connection between emotion and music, emotion model, and music recommendation algorithms.

2.1. *Emotion and music*

Many causal models are proposed to describe the relationship between human emotion and music in psychology research. They are classified as E→M and M→E where E stands for emotion and M stands for music. E→M suggests that the emotion is regarded as the causal one. The emotional experiences of the composer and performer are expressed by musical connotation through the structure and expressive attributes of music[11]. On the other hand, M→E suggests that the music is the causal agent. Listening to music triggers the listener's emotional change and may also cause a listener's physiological response, for instance, increasing the heart rate[4]. Krumhansl's model indicates that the listener's emotional response to music depends on the listener's past emotional experiences[13]. North & Hargreaves[19] showed that a listener's music preference is associated with the affective qualities of a certain context. Music as an emotion-loaded form can convey certain emotion to a listener, and the change in the listener's emotions relates to the listener's previous emotional experiences in a special context.

2.2. *Emotion model*

Generally speaking, there are two kinds of emotion model, the discrete model and the dimensional model. The discrete model suggests that emotion is described by six basic emotions: fear, anger, disgust, happiness, sadness, and surprise[7]. The dimensional emotion theory believes that the emotional categories can be mapped into a bipolar dimensional space. For example, Russell & Steiger proposed that emotion accounts for three independent bipolar dimensions: pleasure-displeasure, arousal-sleepiness, and dominance-submissiveness[20]. Zentner et al proposed a Geneva Emotional Music Scale Model. It includes 33 emotional terms is obtained from a list of adjectives rated by participants[24].

There is no unified model to assign emotion labels to music. Different listeners may get different emotions

from listening to the same music because their distinctive emotion perception of the music. Thus, different emotion categorizations can be obtained from different researches in different domain[12].

2.3. *Music recommendation*

Many music recommender systems recommend music to listeners by tracking the listeners' preference (or listening habits) obtained from their listening history. However, these systems do not consider the listener's context information. The context-aware music recommendation has arisen to offer music personalized to the listener based on the listener's certain context information. Lee[15] proposed an Music-for-My-Mood (MMM) algorithm to recommend music based on context reasoning. Cai et al.[6] used a graphical model to represent the emotional model, in which music is automatically given to the listener when he reads weblogs. Sourcetone[9], a music recommendation system, promises to treat emotional disorders and improve the health of listeners based on their emotional states.

Meanwhile, as described by Yutaka, "This phenomenon by which "similar" people gather is understood as homophily in the context of social network analysis"[18]. In the case of similar taste influenced by each other, a similar listener would listen to similar music in the certain context. The listener will believe his/her trustworthy listeners, who may provide the desired music for him/her in real life.

Emotion-oriented music recommendation is different from general recommendation, in that the listener's preference for music not only depends on both the emotional connotation on the list of music that they have ever enjoyed in the past, and also the listener's music co-rated by the listener and others. In general recommendation, the user's preference merely touches on items co-rated by the user and others. To provide satisfactory music matching his/her emotional state, the emotion-oriented music recommendation not only considers the listener's trust relations to find his/her trustworthy listeners, but also takes into account the listener's preference for emotional connotation of music that is suitable for current emotional state of the active listener.

3. The emotion-oriented music recommendation fusing rating and trust

We present a new approach to match the listener's emotional state with music. The overall architecture of the proposed emotion-oriented music recommendation algorithm is shown in Fig. 1. It contains four parts: the similarity weight, ratings prediction, listener's preference for emotional connotation of music, and candidate set of music. The evolved similarity weight is calculated by the rating value and the trust value based on two thresholds. In particular, we formulate the listener's preference for emotional connotation of music regarding it as the bridge between music and listener, and then combine it with the predicted ratings to produce a candidate set of music.

3.1. *The similarity weight*

This weight measures the similarity of the music preference between two listeners. The greater this weight is, more similar the two listeners are. It is used to predict the ratings of a piece of music that has not been appraised by active listeners. Inspired by trust-based recommendation[17,] we extend the similarity measure with an amalgamation between the rating value and the trust value, and selectively use them to calculate the similarity weight based on two thresholds, under lower threshold and upper threshold, which are determined by the number of music co-rated. The trust value describes the trust degree between two listeners. The rating value is used to weight the correlation between two listeners according to the music that they have both rated. The correlation between two listeners on co-ratings can be calculated by Pearson correlation, Spearman correlation and Cosine correlation. Comparing the existing experimental results, the Pearson correlation is better than the Cosine correlation on calculating the similarity based only on the ratings of the co-rated items[5]. For this reason, we use the Pearson correlation measure to calculate the correlation between two listeners on the co-ratings ($c_{a,b}$), as defined by Eq. (1). $R_{a,i}$ is the rating score that a listener u_a gives to music i , $\overline{R_a}$ is the average rating from u_a , $I_{a,b}$ the co-rated music by u_a and u_b .
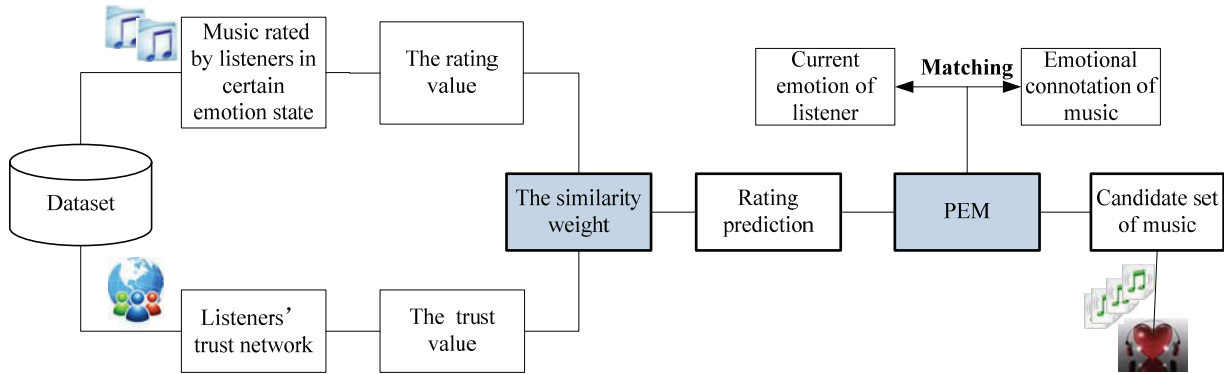
Fig. 1. The architecture of the emotion-oriented music recommendation algorithm.

$$c_{a,b} = \frac{\sum_{i \in I_{a,b}} (R_{a,i} - \overline{R_a}) * (R_{b,i} - \overline{R_b})}{\sqrt{\sum_{i \in I_{a,b}} (R_{a,i} - \overline{R_a})^2} \sqrt{\sum_{i \in I_{a,b}} (R_{b,i} - \overline{R_b})^2}} \quad (1)$$

The similarity weight between two listeners ($sim_{a,b}$) is calculated by Eq. (2).

$$sim_{a,b} = \begin{cases} 0, n \le N_1 \ and \ t_{a,b} = 0 \\ t_{a,b}, n \le N_1 \ and \ t_{a,b} \ne 0 \\ \frac{n}{N_2} \cdot c_{a,b}, N_1 < n < N_2 \ and \ t_{a,b} = 0 \\ \frac{n}{N_2} \cdot c_{a,b} + (1 - \frac{n}{N_2}) \cdot t_{a,b}, N_1 < n < N_2 \ and \ t_{a,b} \ne 0 \\ c_{a,b}, n \ge N_2 \end{cases} \quad (2)$$

Where all symbols and their explanation are shown in Table 1, between $sim_{a,b}$ and $c_{a,b}$ are from range of -1 to 1, and $t_{a,b}$ varies from 0 to 1.

Table 1. Explanation of symbols.

Symbol	Explanation
n	The number of co-ratings
N_1	The lower value of co-ratings
N_2	The upper value of co-ratings
$c_{a,b}$	The correlation between listener u_a and listener u_b based on the co-ratings
$t_{a,b}$	The trust value based on the social network.

When n is less than or equal to N_1, it means the number of co-ratings is small or even absent, so $c_{a,b}$ is unreliable for measuring the correlation between two listeners. The similarity weight is defined by $t_{a,b}$. In the special case when $t_{a,b}$ is not zero. If n is greater than or equal to N_2, the number of co-ratings is enough to guarantee reliability. So $sim_{a,b}$ is defined by $c_{a,b}$. Suppose n is greater than N_1 and less than N_2. This means that the existing number of co-ratings is insufficient to compute the similarity between two listeners; instead, the rating value and the trust value jointly determine the similarity between two listeners. That is, $sim_{a,b}$ is defined by $c_{a,b}$ and $t_{a,b}$.

3.2. Rating prediction

The rating value of music is calculated by Eq. (3).

$$P_{a,j} = \overline{R_a} + \frac{\sum_{b \in U} sim_{a,b} \cdot (R_{b,j} - \overline{R_b})}{\sum_{b \in U} |sim_{a,b}|} \quad (3)$$

Where j is a piece of music for u_a, $\overline{R_a}$ is the average rating from u_a, $R_{b,j}$ is rating that u_b has given to j, and $sim_{a,b}$ is the similarity weight between u_a and u_b.

3.3. Preference for emotional connotation of music

The listener's preference for emotional connotation of music is called PEM for short and defined by the Eq. (4). The PEM measure builds upon the idea of emotional connotation of music to generate the listener's preference that aims at finding the taste of a listener who is in a certain emotional state. N_p and N_n refer to the number of music which is thought to convey the positive and negative emotion, respectively. By mining the listening history of the listeners, we acquire their preference for emotional connotation of music when they were under certain emotions.

$$PEM = \frac{N_p - N_n}{N_p + N_n}.$$ (4)

3.4. *Candidate set of music*

To match a listener's current emotion with music, we obtain the listener's PEM value based on by Eq. (4). Following, we used the predicted ratings and the active listener's PEM to produce the candidate set of music desired by the active listener.

4. Experiments and discussion

We conducted two sets of experiments to evaluate the proposed emotion-oriented music recommendation algorithm. The first group of experiments is employed to evaluate the similarity weight calculated by the rating value and the trust value based on two thresholds without considering the PEM. Two popular datasets are used in this set of experiments. In the second set of experiments, PEM is introduced as the bridge between the music and the listener, to match the listener's emotional state with music on the dataset collected by ourselves in the second set of experiments. The experimental results show that the expanded similarity calculation is helpful in improving the performance of the existing recommendation algorithm, and listener satisfaction increases in the emotion-oriented music recommender system.

4.1. *The similarity weight*

In the first group of experiments, to evaluate the effectiveness of the similarity measure in our algorithm, we compared the other two similarity measures respectively using the correlation on the co-ratings and the trust value on the public datasets named Epinions and Flixster [10]. The correlation on the co-ratings is used to weight user's similarity in the traditional collaborative filtering (CF for short), while the trust value is regarded as user's similarity in the trust-based recommendation (TR for short). The user's similarity is calculated selectively using the correlation on the co-ratings and the trust value based on two thresholds (Eq. (2)) in our proposed algorithm. Without considering the PEM, the proposed algorithm is regarded as a generic recommendation algorithm that would be used in another domain. The recommendation fusing trust and ratings is called RTR for short. It is given in detail as follows:

Input:
R, the dataset of ratings
T, the trust network
u_a, the active user
Output:
m, a list of items recommended for u_a
Steps:
1. Set the matrix of co-rating and the matrix of the trust value of u_a ;
2. For user $u_a \in U$, calculate the similarity weight by Eq. (2);
3. For each of the items, make a prediction by Eq. (3);
4. Find top m predictions items;
5. Return m.

4.1.1. *Dataset*

Two real world data sets, Epinions and Flixster, which were collected from two popular review websites, are used here to verify the performance of RTR.

Epinions is extracted from the trust network datasets released at trustlet.org[16], which was derived from the Epinions.com by Paolo Massa and Paolo Avesani. In the data, the ratings follow the 1, 2, 3, 4, and 5 numerical scales, the trust value is either the value of 0 or the value of 1; obviously, the value of 0 indicates the trust relationship between user A and user B does not exist, while the value of 1 represents that user A trusts user B. This data set contains 49,290 users and 139,738 items. The sparsity of the rating matrices of Epinions is 99.991%.

Flixster is extracted from the social network datasets released by datatang.com, which was derived from Flixster.com in December 2010 by Javier Parra. In this dataset, the friendship of Flixster is used to act as the trust relationship[10], which was weighted at two different levels: 1 and 0. If user B is a friend of user A, the value of 1 represents user A trusts user B in our experiment. There are 1,049,492 users and 66,612 items. The sparsity of the rating matrices of Flixster is 99.991%.

4.1.2. *Evaluation*

Leave-one-out is the most popular technique for evaluating recommender systems. As an offline technique, it uses the recommendation algorithm to predict the hidden rating. Then the product is repeatedly run on all ratings, the predicted rating is compared with the actual rating, and the difference in absolute value

acts as the prediction error. The average of all errors is Mean Absolute Error (MAE). To compensatively measure the error of each user, we adopted the Mean Absolute User Error (MAUE) to evaluate the accuracy of recommendation. MAUE is the AUE's average of all users, which refers to the average of all errors of the special user based on leave-one-out.

If $\{p_1, p_2,...,p_n\}$ refers to the predicted ratings, and $\{r_1, r_2,...,r_n\}$ are the actual ratings, then the MAE is expressed by Eq. (5).

$$MAE = \frac{\sum_{1 \le i \le n} |p_i - r_i|}{n}. \qquad (5)$$

As mentioned above, let U be the set of all users, m refers to the number of users, l_i is the number of ratings given by u_i, then the MAUE and AUE_{u_i} are defined by Eq. (6) and Eq. (7) respectively.

$$MAUE = \frac{\sum_{u_i \in U} AUE_{u_i}}{m}. \qquad (6)$$

$$AUE_{u_i} = \frac{\sum_{1 \le i \le l} |p_i - r_i|}{l_i}. \qquad (7)$$

The coverage is a metric to measure the percentage of predictions given by the recommendation algorithm. It includes the rating coverage and the user coverage, which are respectively denoted as Eq. (8) and Eq. (9).

$$RatingCoverage = \frac{p}{s}. \qquad (8)$$

$$UserCoverage = \frac{k}{m}. \qquad (9)$$

where s is total number of ratings, p is the number of predicted ratings, m is the total number of users, and k is the number of predicted users.

4.1.3. *Experimental results*

Before conducting the similarity measure experiment, we need to firstly determine the lower threshold (N_1) and the upper threshold (N_2) on ratings. In order to determine the appropriate lower/upper threshold, we repeatedly run RTR on Epinions and Flixster. We studied their distribution of items co-rated in Fig. 2. It shows that the majority of users' co-ratings are less than 5.

At first, N_1 is set to 0, while the value of N_2 starts with a value of 5, followed by 10, 20, 30, 40 and 50. If there are more than two co-ratings between listeners, the value of the correlation ($c_{a,b}$) exists in Eq. (2). So the rating coverage and the user coverage have no change when N_2 increases as shown in Table 2 and Table 3. The MAUE and MAE initially decrease and then increase again, reaching a minimum value when N_2 is 20. The results of the experiments show that when N_2 is set to 20, the accuracy of RTR is highest.

Then after we set N_2 to 20, we run a series of experiments to determine N_1. In the two datasets, N_1 is given a value ranging from 0 to 5. As mentioned above, $c_{a,b}$ is on the premise that the number of co-ratings is greater than 2. With N_1 increasing, the users have sufficient co-ratings to enhance the reliability of $c_{a,b}$ in Eq. (2). In the Epinions and Flixster, the accuracy of
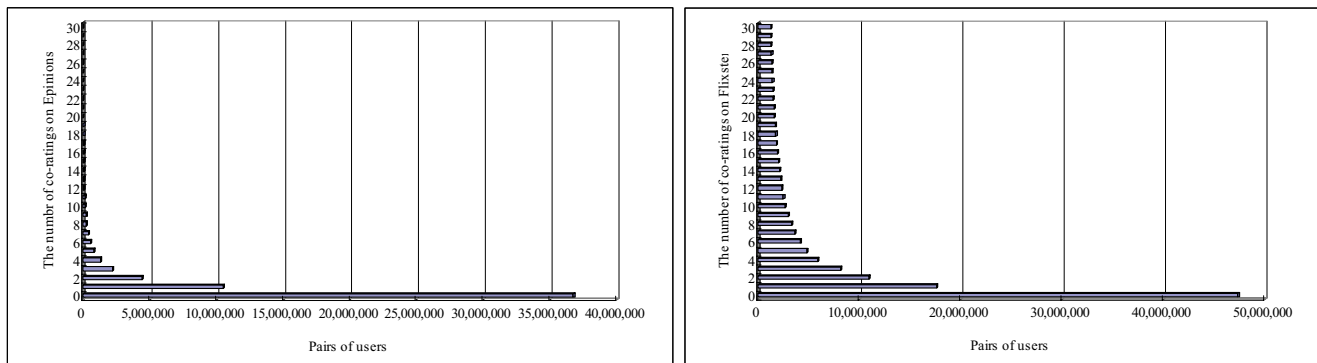

Fig. 2. This is the distribution of items co-rated.

RTR is improved continuously but the coverage decreases as shown in Table 4 and Table 5 respectively.

Based on these analyses, when N_2 equals 20 and N_1 is 0, RTR is most accurate and its coverage is also dramatically improved on Epinions and Flixster. The results of these experiments show that the upper/lower thresholds on Epinions and Flixster keep the same value, this is because two datasets with the same sparsity of the rating matrix.

Table 2. Determine the optimized value of N_2 on Epinions.

N_2	MAE	MAUE	Rating Coverage	User Coverage
5	0.934	1.030	0.583	0.506
10	0.928	1.027	0.583	0.506
20	0.927	1.027	0.583	0.506
30	0.928	1.029	0.583	0.506
40	0.929	1.030	0.583	0.506
50	0.931	1.030	0.583	0.506

Table 3. Determine the optimized value of N_2 on Flixster.

N_2	MAE	MAUE	Rating Coverage	User Coverage
5	0.646	0.747	0.957	0.662
10	0.645	0.746	0.957	0.662
20	0.644	0.746	0.957	0.662
30	0.645	0.749	0.957	0.662
40	0.645	0.750	0.957	0.662
50	0.645	0.751	0.957	0.662

Table 4. Determine the optimized value of N_1 on Epinions.

N_1	MAE	MAUE	Rating Coverage	User Coverage
0	0.927	1.027	0.583	0.506
1	0.927	1.027	0.583	0.506
2	0.921	1.017	0.525	0.443
3	0.916	1.008	0.474	0.408
4	0.910	0.998	0.435	0.388
5	0.908	0.997	0.404	0.378

Table 5. Determine the optimized value of N_1 on Flixster.

N_1	MAE	MAUE	Rating Coverage	User Coverage
0	0.644	0.746	0.957	0.662
1	0.644	0.746	0.957	0.662
2	0.645	0.769	0.953	0.636
3	0.646	0.794	0.948	0.617
4	0.647	0.810	0.943	0.606
5	0.648	0.821	0.940	0.599

To further illustrate our proposed similarity weight, we compared RTR to CF and TR on Epinions and Flixster. The result using Epinions data is shown in Table 6. MAE of RTR is lower than that of CF and higher than that of TR. The rating coverage and the user coverage of RTR are much higher than that of CF and TR, respectively.

Table 6. The results of running CF,TR and RTR on Epinions.

Algorithm	MAE	MAUE	Rating Coverage	User Coverage
CF	0.960	1.069	0.553	0.424
TR	0.904	0.991	0.262	0.355
RTR ($N_1 = 0, N_2 = 20$)	0.927	1.027	0.583	0.506

On Flixster data set, Table 7 shows that RTR and CF have almost the same MAE. Both offer better accuracy than TR. The rating coverage of RTR and CF are much higher than that of TR. The user coverage of RTR is higher than CF and TR.

Table 7. The results of running CF,TR and RTR on Flixster.

Algorithm	MAE	MAUE	Rating Coverage	User Coverage
CF	0.646	0.708	0.947	0.538
TR	0.841	0.929	0.425	0.562
RTR ($N_1 = 0, N_2 = 20$)	0.644	0.746	0.957	0.662

From these experimental results, we can draw the conclusion that RTR is greatly improved on the coverage and also gets a decent accuracy on Epinions and Flixster. That is, the proposed similarity weight using rating and trust has better performance on these two datasets.

4.2. The emotion-oriented music recommendation

In the second set of experiments, PEM is tested as a bridge to match the listener's current emotion with music in the emotion-oriented music recommendation (EMR for short). Following the principle that the categories of emotions are determined by the actual task and domain of research, we consider a scenario where we use music to regulate a listener's negative emotions when they encounter the predicament of emotion in the e-learning process. These negative emotions figured

frequently in e-learning: anxiety, sadness, hopelessness, anger, shame and disgust.

The EMR algorithm is described as:

Input:
R, the dataset of rating
T, the trust network
u_a, the active listener
e_c, the current emotion of u_a

Output:
n, a list of music recommended for u_a

Steps:
1. Generate the matrix of co-ratings on music rated by u_a and the matrix of the trust value of u_a;
2. Predict the ratings by Eq. (2) and Eq. (3);
3. Analyze the PEM of u_a by Eq. (4);
4. Produce n pieces of music after descending order on the predicted ratings, including $\frac{n}{2}*(1+PEM)$ pieces of music inducing positive emotion and $\frac{n}{2}*(1-PEM)$ pieces of music inducing negative emotion;
5. Feedback from the listener satisfaction;
6. Update the ratings and the trust value.

4.2.1. Dataset

There is no popular dataset in the emotion-oriented music recommendation research, because of less available reference data, and lacking of a common categorization system. To illustrate and validate our work, we endeavored to set up a website to acquire data. We named it as "ER" in our experiment.

In order to label the data, volunteers were recruited from Xi'an Jiaotong University in China, 102 volunteers were chosen to rate music at a resource-making website. Participants did not have a history of psychopathology or neurological impairments, as self-reported in an open-ended demographic questionnaire. In this experiment, we collected 1,548 pieces of music taking into account the multidimensional features of songs as well as cross-cultural differences. The music metadata included lyrics text, title, artist, topic, emotion, region, year, etc. The basic metadata relied on the ID3, such as title, artist, and so on.

Music was co-rated by participants in a certain emotional context. Fig. 3 is a sample profile from the ER. All ratings followed the 1-extremely bad, 2-bad, 3-average, 4-good, and 5-perfect numerical scale. While listening, participants could click on the left rectangle to pause/play the song, drag-and-drop the middle rectangle to listen to the song again, drag-and-drop the right rectangle to adjust the volume, and click on the button to modify the annotation. Further, title and artists' names were shown on the webpage. In total, 102 participants gave 21,738 ratings on 1,548 songs. The trust network, established by explicit self-reporting, had a trust rating value of either 0 or 1, where 0 means one listener has no trust in another listener, while 1 indicates complete trust.

4.2.2. Evaluation

A comprehensive method is used to evaluate the performance of EMR with three metrics: MAE, coverage, and listener satisfaction. MAE and coverage as quantitative measures indicates the accuracy and coverage of EMR, respectively. Moreover, since emotion is a subjective psychological experience, self-reporting, originated from feedback in the psychological domain, is an important way of evaluation. Self-report is used to judge listener satisfaction degree (SD for short) in our emotion-oriented music recommendation. It can be expressed by Eq. (10).

$$SD = \frac{\sum_{i=1}^{5} r_i * i}{n}. \tag{10}$$

Fig. 3. This is a sample profile from ER.

where i represents ratings on a scale from 1 to 5 as we have defined, with the higher value representing more satisfaction, r_i is the number of i, and n is the number of the recommended music for a listener.

4.2.3. *Experimental results*

Running the procedure described in 4.1, the best value of N_1 and N_2 was determined on the ER. Focusing on certain emotions, we run EMR on the ratings matrix with a sparsity of 96.663% (and where N_2 =6). Starting with anxiety in a series of experiments, we obtained the results presented in Table 8. EMR and CF both have lower accuracy than TR on MAE and MAUE, EMR is more accurate than CF. Meanwhile, EMR and CF show higher coverage than TR. In particular, we used SD to further illustrate the performance of EMR. Table 9 illustrates the result of listeners' subjective evaluation under the anxiety emotion: EMR gets the highest satisfaction of recommendation. In a word, EMR shows better performance in accuracy, coverage, and listener's satisfaction.

4.3. Discussions

N_1 and N_2 are the key parameters in our algorithm. In order to discover the relationship between the sparsity of rating matrix and the value of N_1 and N_2, we obtain various sparsity of the rating matrices from ER. First, we sorted the listeners by the numbers of ratings they provided, and picked the top 21 most active listeners. Among the music resources, the 147 most frequently rated tracks were selected. Then we selected m listeners from these 21 listeners random, planned to delete the ratings from these m listeners. For each listener, only 3 ratings were retained, and others were removed from the listener's ratings. Running the procedure mentioned above in 4.1, the best N_1 and N_2 was determined from these rating matrices. The final results are presented in Table 10. It shows that sparsity increases when N_2 increases. We can say that the sparsity of the rating matrix determines the value of N_2.

Table 10, shows that N_1 is usually fixed at a value of 0, and the value of N_2 is correlated with the sparsity of the rating matrix. Next, we try to establish a general model to ascertain the best value of N_2 on the rating

Table 8. The results of running CF, TR and EMR.

Evaluation	Algorithm	Certain Emotion Context					
		Anxiety	Anger	Disgust	Sad	Shame	Hopeless
MAE	CF	0.872	0.872	0.826	0.902	0.809	0.932
	TR	0.821	0.817	0.802	0.888	0.776	0.88
	EMR	0.870	0.866	0.826	0.900	0.810	0.933
MAUE	CF	0.960	0.968	0.901	0.952	0.872	0.999
	TR	0.827	0.862	0.825	0.971	0.819	0.937
	EMR	0.938	0.951	0.878	0.948	0.855	1.010
Rating Coverage	CF	0.737	0.737	0.727	0.745	0.724	0.742
	TR	0.634	0.623	0.626	0.625	0.607	0.625
	EMR	0.742	0.742	0.732	0.750	0.730	0.747
User Coverage	CF	0.578	0.640	0.662	0.620	0.680	0.667
	TR	0.470	0.507	0.514	0.514	0.514	0.494
	EMR	0.578	0.653	0.662	0.620	0.681	0.681

Table 9. Listener's satisfaction degree in the anxiety emotion.

Algorithm	The number of ratings					SD
	1	2	3	4	5	
CF	1	9	27	46	9	3.589
TR	1	8	29	35	17	3.656
RTR	1	9	28	35	17	3.644
EMR	0	8	18	45	19	3.833

matrix. The data originates from Table 10, and is represented by Table 11. According to the Pearson correlation test, the value of correlation between N_2 and sparsity of the rating matrix is 0.936. According to the Growth Curve of N_2 in Fig. 4, the test of goodness of fit, the test of F and the test of significance (Table 12) all show that the model is feasible. The value of N_2 can be

predicted by sparsity of the rating matrix using Eq. (11) from Fig. 4.

$$y = e^{-1.87 + 3.85x}. \qquad (11)$$

Table 10. Determine the optimal value of N_1 and N_2.

m	Sparsity of rating matrix	N_1	N_2
1	0.464	0	1
10	0.7154	0	2
14	0.8364	0	4
17	0.8966	0	5
19	0.9796	0	7

Table 11. N_2 and sparsity of the rating matrix.

Sparsity	0.4640	0.7154	0.8364	0.8966	0.9796
N_2	1	2	4	5	7

Table 12. The test value and parameters.

The	The test value			Parameter	
Growth	R2	F	Sig	b0	b1
Curve	.980	150.197	.001	-1.870	3.850

Where the horizontal axis represents the sparsity of the rating matrix and vertical axis represents the best value of N_2.

The results of the above-mentioned experiments show that the sparsity of the rating matrix has a major impact on choosing what data to be used to calculate the similarity weight, the trust value or correlation to the rating value.

5. Conclusions and Future work

By integrating the psychological theories with the existing recommender technique, we have proposed an emotion-oriented music recommendation algorithm. Compared to traditional collaborative filtering and trust-aware recommendation, our algorithm adopts the rating and trust as new features to measure the listener similarity. The emotional connotation of music is suited to the current emotion of the listener based on PEM from his/her listening history. Meanwhile, a comprehensive measure is presented to evaluate the performance of the algorithm by the metrics including MAE, MAUE, coverage, and listener satisfaction.

The main contributions of this paper are threefold:

First, the similarity weight is calculated to selectively use the trust value and the correlation on co-ratings
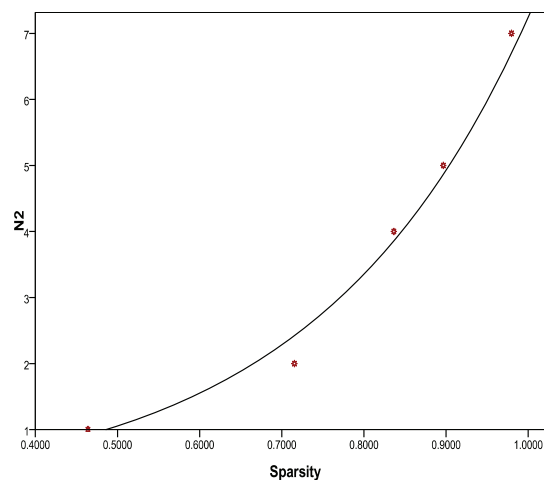

Fig. 4. Integrated fitting curve for experiment data.

based on two thresholds determined by the sparsity of the rating matrix. Our proposed similarity measure differs from both the correlation between two users relying on co-rating in traditional collaborative filtering, and from the trust value of trust-aware recommendation.

Second, we have proposed an emotion-oriented music recommendation algorithm based on PEM to recommend the appropriate music matching the listeners' emotional state. Our algorithm not only takes into account music predicted by other listeners' behavior, but also considers the emotional connotation of music rated by the active listener himself in a certain emotional context.

Third, we have explored a set of comprehensive metrics to evaluate our emotion-oriented music recommendation algorithm. The predictive accuracy of the algorithm by MAE and the satisfaction of listeners measured by self-report can fairly evaluate the performance of emotion-oriented music recommendation.

The experimental results show that the emotion-oriented music recommendation algorithm outperforms the traditional collaborative filtering on predicted accuracy, and outperforms the trust-aware recommendation on coverage as well. The listener is satisfied with the music suggested by the emotion-oriented music recommendation algorithm.

In the future work, the trust network is part of an ongoing work on our proposed algorithm. One is to enlarge the range of the trust value with respect to the range of Pearson's correlation and the similarity weight. Another will focus on the distrust inspired by the

literature[23] which demonstrated that the distrust is as important as the trust in the computable model of trust in recommendation.

Acknowledgements

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(6) (2005) 734-749.
2. G. Adomavicius and A. Tuzhilin, Context-aware recommender systems, *Recommender Systems Handbook*, (2011). 217-253.
3. I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose, Enriching user profiling with affective features for the improvement of a multimodal recommender system, In *Proceedings of the ACM International Conference on Image and Video Retrieval* (2009) , pp. 29:1--29:8.
4. D. L. Bartlett, Physiological responses to music and sound stimuli, *Handbook of music psychology*, 2 (ACM, 1996) 343-385.
5. J. S. Breese, D. Heckerman, and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, (Morgan Kaufmann, 1998), pp. 43-52.
6. R. Cai, C. Zhang, C. Wang, L. Zhang, and W. Y. Ma, Musicsense: contextual music recommendation using emotional allocation modeling, In *Proceedings of the 15th international conference on Multimedia* (ACM, 2007), pp. 553-556.
7. P. Ekman, An argument for basic emotions, *Cognition & Emotion*, 6(3-4) (1992) 169-200.
8. J. Herlocker, J. A. Konstan, and J. Riedl, (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms, *Information retrieval*, 5(4) (2002) 287-310.
9. A. Huq, J. P. Bello, and R. Rowe, Automated music emotion recognition: A systematic evaluation, *Journal of New Music Research*, 39(3) (2010) 227-244.
10. M. Jamali, and M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, In *Proceedings of the fourth ACM conference on Recommender systems* (ACM, 2010), pp. 135-142.
11. P. N. Juslin, and J. Sloboda, *Handbook of Music and Emotion: Theory, Research, Applications*, (2009) OUP Oxford.
12. M. Kaminskas, and F. Ricci, Contextual music information retrieval and recommendation: State of the art and challenges, *Computer Science Review*, 6(2-3) (2012) , 89-119.
13. V. J. Konečni, A. Brown, and R. A. Wanic, Comparative effects of music and recalled life-events on emotional state, *Psychology of music*, 36(3) (2008) 289-308.
14. I. Konstas, V. Stathopoulos, and J. M. Jose, On social networks and collaborative recommendation, In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (ACM, 2009), pp. 195-202.
15. J. Lee, Music for my mood: A music recommendation system based on context reasoning, *Smart sensing and context*, LNCS 4272, (2006) 190-203.
16. P. Massa, and P. Avesani, Trust-aware bootstrapping of recommender systems, In *ECAI 2006 Workshop on Recommender Systems*, (Riva del Garda, Italy,2006), pp. 29-33.
17. P. Massa, and P. Avesani, Trust metrics in recommender systems, *Computing with social trust*, (2009) 259-285.
18. Y. Matsuo, and H. Yamamoto, Community gravity: measuring bidirectional effects by trust and rating on online social networks, In *Proceedings of the 18th international conference on World wide web* (ACM, 2009), pp. 751-760.
19. A. C. North, and D. J. Hargreaves, Situational influences on reported musical preference, *Psychomusicology: Music, Mind & Brain*, 15(1) (1996) 30-45.
20. J. A. Russell, and J. H. Steiger, The structure in persons' implicit taxonomy of emotions, *Journal of Research in Personality*, 16(4) (1982) 447-469.
21. S. Saarikallio, and J. Erkkilä, The role of music in adolescents' mood regulation, *Psychology of music*, 35(1) (2007) 88-109.
22. U. Shardanand, and P. Maes, Social information filtering: algorithms for automating "word of mouth", In *Proceedings of the SIGCHI conference on Human factors in computing systems* (ACM Press/Addison-Wesley Publishing Co,1995), pp. 210-217.
23. P. Victor, C. Cornelis, M. De Cock, and P. Pinheiro da Silva, Gradual trust and distrust in recommender systems, *Fuzzy Sets and Systems,* 160(10) (2009) 1367-1382.
24. M. Zentner, D. Grandjean, and K. R. Scherer, Emotions evoked by the sound of music: characterization, classification, and measurement, *Emotion*, 8(4) (2008) 494.