International Journal of u- and e- Service, Science and Technology
Vol.9, No. 9 (2016), pp.101-110
http://dx.doi.org/10.14257/ijunesst.2016.9.9.11

DNALS: A Recommendation Algorithm Based on Chinese Vocabulary Emotion Analysis of Songs

Chengzhou Fu, Weiquan Zeng,Yong Tang*,Lingxiao Chen and Jingmin Wei

School of Computer Science, South China Normal University, Guangzhou
*fucz@m.scnu.edu.cn,*ytang@scnu.edu.cn*

Abstract

The lyrics play important roles in emotion expression of songs, as well as the emotion words of the lyrics reveal the emotion theme of songs. By extracting the lyrics' emotion information, this paper uses songs' emotion themes as the recommended standard. We get all kinds of emotion information extraction coefficient combining emotion gene sequences of lyrics which named DNALS (DNA for Lyrics of Songs), and then put forward the DNALS recommendation algorithm based on emotion analysis. By analyzing the emotion information of user's historical music data, we recommend to the user a list of songs with similar themes emotion, so as to help the users to find songs for mood.

Keywords*: Emotion Gene; Vocabulary Emotion; Hadoop; Recommendation Algorithm*

1. Introduction

Songs play an important role in people's daily life, and those are important tools for people to cultivate their emotions. The Internet gradually transformed into multimedia storage, sharing and distribution center from multimedia communication tools. According to statistics, the total amount of Chinese music has reached million levels. In the case of a large expansion of this information, for Internet users, it takes a lot of time and energy to search for useful information, and it is more difficult to retrieve songs that are in line with their own tastes. Currently the mainstream online music platforms have launched songs' recommendation, such as Kugou, Netease Cloud Music, Xiami Music, and Douban FM. They calculate the user's historical music data based on the analysis of the user's behavior, so as to select a group of songs for users to achieve the purpose of the recommendation.

Recommendation system is widely used in E-commerce, social networks, and other fields, as it shortens the time of user choice, improves the user experience as well as brings the actual economic benefits directly to the service providers. Consequently, the research on the recommendation system is a very meaningful thing.

A practical recommendation system should have the following characteristics:(1) Automatically add a track to meet the conditions;(2) The speed of the algorithm meets the requirements of the user experience;(3) Most of the recommended results are accepted by users, and the precision and recall are within the threshold range required.

To achieve these three goals, we need to pay attention to the following contents of studying the recommended algorithm:(1) The study of the music itself;(2) Research on the user's preferences;(3) A study of the relationship between music and the user.

The main work of this paper is to draw the emotion information about songs, and combine the ICTCLAS System (Institute of Computing Technology, Chinese Lexical Analysis System) developed by the Chinese Academy of Sciences (CAS) [1], extract words and attributes, statistic emotion vocabulary in the lyrics, quantify the emotion factors by extract DNALS of the song, query the list of songs with similar emotion from the song data sets to give recommendation.

*Corresponding Author

ISSN: 2005-4246 IJUNESST

In order to tap the emotion information which contained in songs, we calculate the emotion genes of all the songs in data sets. This complex operation can be done in a way that is offline. The result of the operation is stored in the database to be accessible in query. In this paper, experiments show that the precision and recall of the algorithm proposed are relatively ideal.

2. Related Work

In recent years, with the development of Web 2.0 and digital music technology, the research of the song recommendation algorithm is concerned by the scholars. Song recommendation technology provides users with songs that are in line with their interests according to the content features of music and users' behavior. This technology can quickly and accurately retrieve the list of songs that fit their taste for different users, even find out the music interests that they have not found by themselves. The way of accessing the music resource has a fundamental change, namely transit to system automation recommendation from artificial retrieval. The song recommendation method as recommendation based on the contents of, collaborative filtering recommendation, recommendation based on association rules and knowledge based recommendation.

The Ringo system was completed in the year 1995 [2], marking the birth of the first music recommendation system in history. Ringo system can not only recommend music that users like, and predict that music they do not like, but also predict the user's score for a particular music. After a long time, music recommendation is generated through the basic information of the song, such as composers, singers and songs, etc. In this way, the result is lack of pertinence, and it also appears to be monotonous. The paper [3] proposes a scheme for recommending music automatically by considering both personal and general musical predilections, and for blending such music into a mixed clip for seamless playback. For music recommendations, they first analyze social networks to identify a general predilection for certain music genres that depends on time and location. Songs that are generally preferred within a certain time period and location are identified through statistical analysis.

The paper [4] expands the semantic lexical sets of typical features of songs by analyzing the strong association rules between the data sets. A recommendation list is obtained according to the correlation degree between the song and the semantic vocabulary. Reduce the dimension of the data sets by the algorithm of SVD (Singular Value Decomposition) and find the feature information which can represent the song in the low dimensional semantic space, using these features to calculate the similarity between songs to get the song list, to achieve the purpose of the recommendation. The paper [5] analyzes the relationship among users, tag and item in the music social network Last.fm based on graph node structure similarity and random walk with restart model, then, construct the relationship among music and labels. After getting a preliminary list of music recommendations and the indirect association of music, get the recommended list by reordering according to the proposed algorithm fusion results, so as to realize the music recommendation algorithm.

The paper [6] draws the audio features of the songs and the social tags through data mining and digital signal processing methods and constructs the feature database of song. Then, pretreat song features and reduce the dimension of the feature vector showing the correlation between songs in the two-dimensional music space. The paper gets a large number of actual user's tags from the commercial music website, analysis 80 audio features of songs, constructs a music feature database, and realizes the music recommendation prototype system.

The paper [7] proposes a personalized music recommendation algorithm based on tags and ratings (P-RABOTR). Establish a data association diagram between user, tag and music through the user's music and label operation and get a low dimensional space to

maintain the relationship between the three types of data by dimension reduction. Further, recommend the nearest music for the target users according to the mapping coordinates and the Euclidean distance between the user and the music in this low dimensional space. To do further abstraction of the algorithm model as narrated above, a large scale music recommendation algorithm (LS-RABOR) based on score is obtained. The algorithm builds a hierarchical model of four layers only by the user's music score, targeting users as the starting point to value transfer to other nodes. Finally, provide users with the recommended results according to the value of music.

Our study analyzes the emotion theme of songs from the perspective of emotion based on lyrics information, and then, according to the user's history, to recommend to users the songs list with roughly similar emotion.

3. Emotion Vocabulary Processing

3.1. Emotion Vocabulary Classification

In this paper, we extract the words with emotion significance in the vocabulary, and classify emotion words, taking the characteristics of Chinese language into consideration. 14 kinds of emotions are summarized, as shown in Table 1.

Table 1. Vocabulary Emotion Classification

Type	Symbol	Type	Symbol	Type	Symbol
喜(fund)	fn	恐(frightened)	ft	敬(respected)	rp
怒(angry)	an	惜(pity)	pt	平静(calm)	cl
哀(sorrow)	sr	恨(hateful)	ht	失望(disappointed)	da
乐(happy)	hp	惊(scared)	sc	激动(excited)	ex
愁(sad)	sd	思(yearning)	yn	—	—

Through further analysis, the fourteen kinds of emotions can be divided into positive (as shown in Table 2), neutral (as shown in Table 3), and negative (as shown in Table 4). We split the lyrics next based on the emotion pool above. In the process of extracting the words, we find that the words with negative relationship and double-negative relationship will distract the judgment of emotion. Therefore, in order to eliminate the interference, we put forward the filtering method of "negative" and "double-negative".

Table 2. Emotion Vocabulary (Positive, a Total of 124)

Type	Vocabulary
rp	称赞,恭敬,敬慕,敬佩,敬畏,敬仰,敬重,钦敬,思慕,羡慕,向往,仰慕, ...,尊重
cl	安定,安宁,安然,安稳,安详,从容不迫,快慰,冷静,满意,宁静,惬意, ...,自在
ex	百感交集,悲愤填膺,悲喜交集,动人心弦,感人肺腑,激昂慷慨, ...,心潮澎湃
fn	畅快,称心,称心如意,踌躇满志,春风得意,大喜过望,得意忘形, ...,愉快
hp	闷闷不乐,鼓乐齐鸣,欢快,欢乐,欢喜,津津乐道,钧天广乐, ...,自得其乐

Table 3. Emotion Vocabulary (Neutral, a Total of 17)

Type	Vocabulary
sc	诧异,吃惊,好奇,惊奇,惊讶,惊疑,震惊
yn	痴心妄想,驰思遐想,挂念,胡思乱想,怀念,苦思冥想,想奇思妙, ...,想念

Table 4. Emotion Vocabulary (Negative, a Total of 200)

Type	Vocabulary
an	暴跳如雷,勃然变色,勃然大怒,大发雷霆,大怒,发怒,发指眦裂, ...,义愤填膺
da	黯然销魂,耻辱,垂头丧气,大失所望,丢丑,丢人,孤独,孤寂,害臊, ...,自暴自弃
sr	哀毁骨立,哀戚,哀伤,哀声叹气,哀痛,哀痛欲绝,哀怨,悲哀,悲愤, ...,心如刀割
ft	不安,不知所措,胆怯,胆小,度日如年,发憷,发慌,方寸大乱,害怕, ...,坐立不安
ht	怅恨,仇恨,仇怨,敌视,敌意,妒忌,反感,愤恨,愤慨,愤懑,嫉妒,可恶, ...,憎恨

Type	Vocabulary
pt	懊悔,抱歉,忏悔,垂头丧气,顿足捶胸,后悔,悔悟,可怜,可惜,内疚, ...,长吁短叹
sd	黯淡,黯然失神,懊恼,不甘,愁眉不展,愁眉蹙额,愁眉紧缩,愁眉苦脸, ...,郁郁寡欢

3.2. Interference Information Filtering

We divide negative sentences into two categories: explicit negation and implicit negation. The former uses negative terms such as "不(no, not)", "没(without)", and "无(without)", which are relative to the dominant structure, refers to the negative meaning without negation form, largely from the perspective of pragmatic effect. Owing to the implicit denial beyond the research category, our experiment deals only with the explicit structure of negation, that is, the negative relationship expressed by negative words as shown in Table 5.

Table 5. Negative Words of Emotion Vocabulary

Negative words
不,不是,没,没有,无,否,非...

In Chinese grammar, there is not only a negative relationship, but also a double-negative. In order to achieve better results, we consider the double-negative situation as the same as affirmation to eliminate it. In the process of extracting the emotion information of the lyrics, we filter the double-negative modified words, and during operation period, we converse the double-negative for positive emotions, and make it more strongly than the general narrative in the degree of emotion. So in the experiment of this paper, we enhance the weight value of double-negative words (weight=2). The double-negative words as shown in Table 6.

Table 6. Emotion Vocabulary Double-Negative

Double-negative words
决非,并非,不是不,并不是不,不可能不,不...不...,没有...不...,非...不...,非...不可...,无不,无非,不无,未必不...,不得不...,不能不,不会不, ...,不可不

Assuming that lyrics are processed by word segmentation, a total of emotion words is *N*, the number of emotion words in the general narrative is *General*, the emotion vocabulary Negative with negative relationship is *Double*. The calculation method of emotion factors *EmotionCount* is as follow:

$$EmotionCount = \sum_{n=1}^{N}\left[\mathrm{Bool}(word_n, General) - \mathrm{Bool}(word_n, Negative) + weight \times \mathrm{Bool}(word_n, Double)\right] \quad (1)$$

Due to the analysis, when *EmotionCount*<0, we set the *EmotionCount* to zero. The calculation function of the *Bool* above expresses as follow:

$$\mathrm{Bool}(\mathrm{word}_n, Emotion) = \begin{cases} 1 & word_n \in Emotion \\ 0 & word_n \notin Emotion \end{cases} \quad (2)$$

In which, word_n is the No.*n* emotion vocabulary and *Emotion* is the emotion type in the lyrics text.

4. Algorithm Implementation

4.1. DNALS Model

As shown in Figure 1, according to the last section described, we divide the word into fourteen kinds of emotion such as fn (fun), an (angry), sr (sorrow), .., and ex (excited),

and give each emotion 4 bits to store the number of words of a kind of emotion extracted from the lyrics.

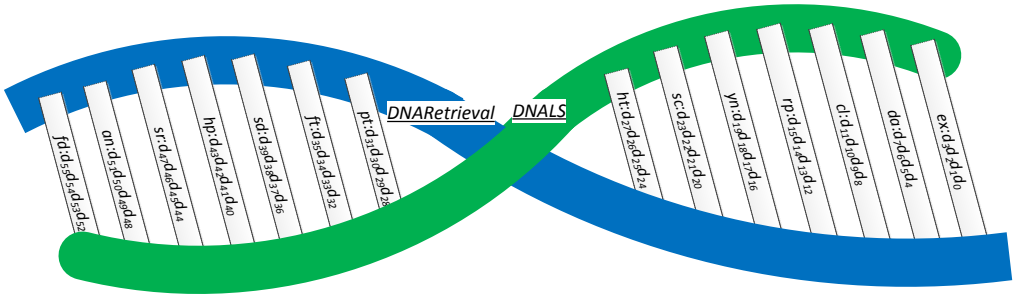

Figure 1. DNALS Model

As shown in Figure 2, due to our experiment, we find that 0.9999 of the count values are not more than 15 for single type of emotion. When the number n>15, we can set the binary value to $(1111)_2$. In this way, there will be a total of 56 bits (4 bits×14) as $d_{55}d_{54}d_{53}\ldots d_2d_1d_0$ that we call the emotion gene sequence. The emotion gene sequence is the genetic factor in the biological technology, the main material of genetic variation, and the basic structure and performance of creature. Similarly in this paper, emotion words in the lyrics are known as emotion genes (DNALS), and we call every kind of emotion in the gene as factor (*DNAFactor*). The calculation results between DNALS constitute motional gene pool (*DNABase*) in which the emotion gene sequence (*DNARetrieval*) is used to retrieve the recommended list of feelings when *DNARetrieval* and *DNALS* for equivalent matching.

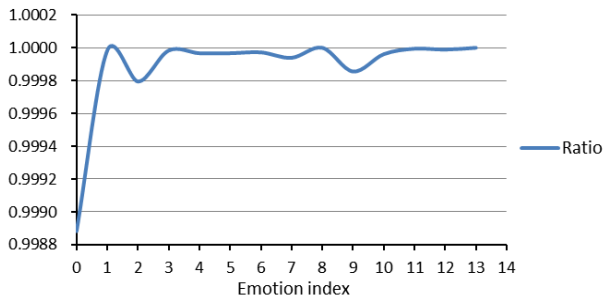

Figure 2. Emotion Count Experiment

4.2. DNALS Data Sets Training

Definition (DNALS @ Operation). Suppose that there are *DNALS*s of *N* Songs. We carry out "@ Operation" on the numerical value of emotion factor for class *i* in the emotion gene pool *DNABase*, calculated as follows:

$$DNAFactor_i = \left\lfloor \left(\sum_{n=1}^{N} DNAFactor_{(n,i)} \right) \Big/ N \right\rfloor$$

$DNAFactor_i$ is the emotion gene *i*, $DNAFactor_{(n,i)}$ is the emotion gene *i* of $DNALS_n$.

We need to train the song genes of the audience before recommending. Assuming that the user *A* has listened to *N* songs, and the *DNALSs* of these songs have been calculated. Extract the values not equal to zero, respectively, $DNAVal_0$, $DNAVal_1$, …, $DNAVal_n$. At this point, we

construct the gene pool of user A by using these song samples. The method expression becomes:

$$DNABase = DNAVal_0 @ DNAVal_1 @ DNAVal_2 @ ... @ DNAVal_n \quad (3)$$

4.3. DNALS Recommendation Algorithm

According to the formula (4), when the *DNABase* is not zero, the track training data sets of calculation results is effective. In this case, *DNARetrieval=DNABase*, that is the so-called convergence type recommended mode.

$$DNABase = DNAVal_0 @ DNAVal_1 @ DNAVal_2 @ ... @ DNAVal_n \neq 0 \quad (4)$$

In formula (5), when *DNABase* is zero, the result of the training set of the tracks is invalid. We need to re-calculate the training set, that is, the broad generic recommendation model.

$$DNABase = DNAVal_0 @ DNAVal_1 @ DNAVal_2 @ ... @ DNAVal_n = 0 \quad (5)$$

Under these circumstances, we have to split the fourteen kinds of emotion again, and the emotion factor of the gene sequence is as the following processing:

Let $DNAValFactor(n,i)$ be the $DNAVal_n$ of the type i, then the *DNABase* of the N songs of the type i of emotion factors can be expressed as follows:

$$DNAFactor_i = \sum_{n=1}^{N} DNAValFactor_{(n,i)} \qquad i = 0,1,2,...,13 \quad (6)$$

In the equation (6), $DNAFactor_i$ represents the binary value of emotion factor i. Through the calculation, we get the numerical value of all kinds of emotions in *DNABase* such as $DNAFactor_0$, $DNAFactor_1$, ..., $DNAFactor_{13}$. Let $DNAFactor_0'$, $DNAFactor_1'$, ..., $DNAFactor_k'$ be the k values which are not zero among the $DNAFactor_0$, $DNAFactor_1$,...,$DNAFactor_{13}$. In order to make numerical value of each operation of the algorithm has comparability, we will carry on standardized treatment on numerical values. In this case, $DNAMinFactor = min(DNAFactor_0', DNAFactor_1', ..., DNAFactor_k')$, and the standard treatment method for all kinds of emotion factor in *DNABase* as follows:

$$DNABaseFactor_i = \lfloor DNAFactor_i / DNAMinFactor \rfloor \qquad i = 0,1,2,...,13 \quad (7)$$

Finally, we obtain the retrieval gene for the broad model recommendation is as: $DNARetrieval = DNABaseFactor_{13}DNABaseFactor_{12}\ldots DNABaseFactor_1DNABaseFactor_0$.

Example: On the assumption that User A listened to three songs. The *DNALS*s of the songs are $DNAVal_0$, $DNAVal_1$, and $DNAVal_2$, respectively, as shown in Table 7.

Table 7. DNALS of User *A*

DNALS	Value
$DNAVal_0$	0010 0010 0000 0010 0010 0010 0000 0000 0000 0000 0000 0100 0000 0101
$DNAVal_1$	0011 0010 0000 0000 0010 0010 0000 0000 0000 0000 0000 0100 0000 0100
$DNAVal_2$	0010 0000 0010 0010 0010 0010 0000 0000 0000 0000 0000 0100 0000 0001

Hypothetically, the corresponding emotion x's numerical value of $DNAVal_0$,$DNAVal_1$, $DNAVal_2$ respectively are $x_0 = (0010)_2$, $x_1 = (0011)_2$, $x_2 = (0010)_2$, and the result of "DNALS@Operation" is as follows:

$$DNABaseFactor = \lfloor (x_1 + x_2 + x_3)/3 \rfloor = \lfloor (2+3+2)/3 \rfloor = 2$$

Assuming that the DNALS seed value of user A is expressed as $DNABase_A$, the A's "DNALS@Operation" can be expressed as follows:

$$DNABase_A = \begin{array}{r} 0010\,0010\,0000\,0010\,0010\,0010\,0000\,0000\,0000\,0000\,0000\,0100\,0000\,0101 \\ @\,0011\,0010\,0000\,0000\,0010\,0010\,0000\,0000\,0000\,0000\,0000\,0100\,0000\,0100 \\ @\,0010\,0000\,0010\,0010\,0010\,0010\,0000\,0000\,0000\,0000\,0000\,0100\,0000\,0001 \\ \hline 0010\,0001\,0000\,0001\,0010\,0010\,0000\,0000\,0000\,0000\,0000\,0100\,0000\,0011 \end{array} \quad (8)$$

Due to $DNABase_A \neq 0$, the $DNARetrieval_A$=0010 0001 0000 00010010 0010 0000 0000 0000 0000 0000 0100 0000 0011.The process of the recommendation algorithm is shown in Figure 3, and the pseudo code of the algorithm is given as follows:

Algorithm:	DNALS Songs recommendation
Data:	Lyrics data sets **lyrics[]**
Result:	Recommended repertoire **songs_index[]**
Step 1:	Chinese word segmentation operation **segmentation()**;
Step 2:	Extract emotion vocabulary **find_emotion_words()**, returns the corresponding vocabulary and the emotion type **emotion_kind** and count **emotion_count**;
Step 3:	Computational gene pool **calculation_dna_base()**, when the return value is not zero, goto Step 3.1, otherwise goto Step 3.2; **Step 3.1:** Invoke convergent recommendation **convergent_recommendation()**, goto Step 4; **Step 3.2:** Invoke Broad recommendation **broad_recommendation()**, goto Step 4;
Step 4:	Search for appropriate recommendations based on gene pool **find_index_in_song_libs()**, return the number of songs;
Step 5:	Put the number of songs found in the array **put_in(songs_index[i++])**, repeat Step 4 until all the songs are traversed;
Step 6:	Return the number of all recommended tracks **songs_index[]**.

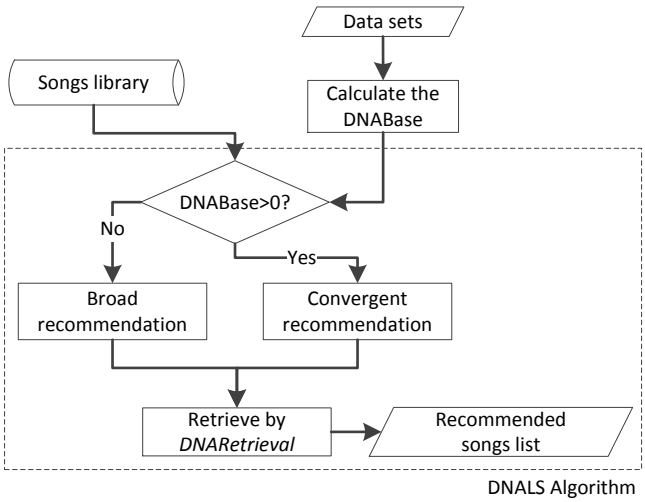

Figure 3. Recommendation Algorithm Process of DNALS

5. Simulation Experiment and Analysis

The lyrics data used in this experiment are derived from the Baidu Inc. opening interface. Through processing and filtering, we have grabbed 182088 lyrics data totally. The configuration of the experimental equipment is as follows: Debian 5.0, Intel (R) Core (TM) CPU i7-2600 @ 3.4 GHz, RAM 8.0 GB, and the lyrics data sets storage in MySQL database.

We filter the collected songs, and set all song data sets for $U=\{u_1, u_2, \ldots, u_n\}$, the recommended data sets is $T=\{t_1, t_2, \ldots, t_n\}$, and the relationship between U and T is $T \subseteq U$, in which any element of T satisfies $\forall t \in U$, and $\forall t(t \neq 0)$.

In order to improve the efficiency of data processing, we use Hadoop distributed processing system to work out the lyrics data sets. Furthermore, we design the key-value pair as :< “id-emotion”, count>, which *id* presents the index of songs; *emotion* presents the type of emotion; and *count* presents the total of emotion depend on the processing. The MapReduce process is shown as Figure 4.

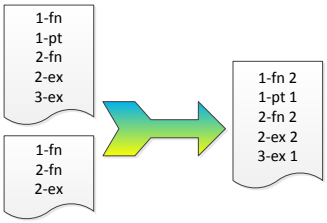

Figure 4. MapReduce Batch Process

As shown in Table 8, in the process of simulation experiment, we randomly select three songs from the data sets, and extract emotion factors to the lyrics of these songs. Those words with "__" are considered as the emotion words, which with "‿" are negative words, and other with "…" are double-negative words. Assuming that user *A* has listened to these three songs, we can use user's initial data sets to build the user's emotion gene pool, and further calculate its *DNABase*. In the following, we extract the emotion gene database of all the song lyrics, and part of the results obtained from the song DNALS construction is as shown in Table 9.

Table 8. Songs Emotion Factor Extraction Results

Title	General	Lyrics	Negative	Double	Lyrics
会呼吸的痛	sr	没看你脸上，张扬过哀伤	—	—	—
	sr	我终于到达，但却更悲伤	—	—	—
	pt	后悔不贴心会痛	—	—	—
	pt	后悔不贴心会痛	—	—	—
	mf	想念是会呼吸的痛	—	—	—
	mf	想念是会呼吸的痛	—	—	—
我不会喜欢你	fn	所以我让自己那么喜欢你	fn	—	我不会喜欢你
	fn	我喜欢了，我讨厌了	fn	—	我必须说我真的不会喜欢你
	ht	我想我讨厌，讨厌骄傲的你	fn	—	我不喜欢你占据我所有思绪
	ht	也讨厌美好，美好的那个你	fn	—	别笑了，别笑了，我不会喜欢你
	ht	于是我要自己假装讨厌你	fn	—	你不必懂，我真的不会喜欢你
	ht	我喜欢了，我讨厌了	fn	—	别想了，别想了，我不会喜欢你
	—	—	fn	—	别想了，别想了，我不会喜欢你
躲进爱里面	sd	总担心自己走不进	—	fn	不是不喜欢你

Table 9. Songs Gene Construction Results

Index	Title	DNALS
2433	自由自在	0011 0000 0001 0000 0000 0000 0000 0000 0000 0000 0000 0011 0000 0000
2439	心里有数	0100 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
5124	会呼吸的痛	0000 0000 0010 0000 0000 0000 0010 0000 0000 0010 0000 0000 0000 0000
141120	不要说话	0001 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
891465	红玫瑰	0010 0000 0000 0000 0000 0010 0000 0000 0000 0000 0000 0000 0000 0000
1057438	我不会喜欢你	0010 0000 0000 0000 0000 0000 0000 0101 0000 0000 0000 0000 0000 0000
1068086	倒不如	0110 0000 0000 0000 0000 0000 0010 0000 0000 0000 0000 0000 0001 0000
1070771	躲进爱里面	0011 0000 0000 0000 0001 0000 0000 0000 0000 0000 0000 0000 0000 0000

After the previous experiments, we query the tracks meeting the conditions of the algorithm using the constructed gene pool. If we query all the songs for each user in recommendation system, the recommendation algorithm will be inefficiency because of huge amount of computation. To solve this problem, we use the Top-*k* method to improve efficiency and reduce resource consumption. Stop a current query until getting the track to achieve the preset *k* value, and record the index value at this time to facilitate retrieve. The results of the three groups are shown in Table 10.

Table 10. Songs Recommended Experimental Results

Index	Initial list	Recommended list
1	我不会喜欢你,彩虹天堂,小情歌	我就是爱你,祝你幸福,专属制作,红地毡,受之有愧,吻所有女孩,阅读过去,爱我不一样,80 年代,完美的演员,集体鼓舞,宠物,Another Me,只恨我一个人
2	漂洋过海来看你,讲不出再见,小酒窝	渲染离别,说一声再见,爱情是圆的,向家驹叹息,归途,代替爱,就让这首歌,海浪泪痕,说再见,飘零的艺人,宠坏,慢慢等
3	红玫瑰,白玫瑰,淘汰	爱趁现在,兄弟,最佳努力奖,从今日开始,两粒糖,心上,如果我会功夫

Let C_{like_recom} for the total of songs that users actually like in the recommended results, C_{sum_recom} for the total of songs algorithm recommended, C_{like_lib} for the total of songs users like in music data sets. The R (Recall) of a recommendation algorithm indicates the probability that a user like the recommended song, as shown in formula (9). The P (Precision) indicates the probability that the user actually likes the recommended songs, as shown in formula (10). The precision and recall of the results are shown in Figure 5.

$$R=\frac{C_{like\_recom}}{C_{like\_lib}}\times 100\% \tag{9}$$

$$P=\frac{C_{like\_recom}}{C_{sum\_recom}}\times 100\% \tag{10}$$

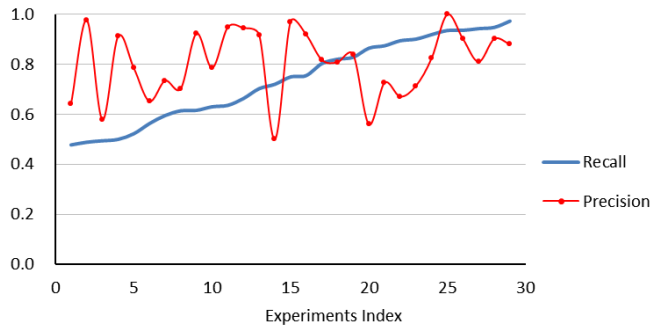

Figure 5. The Precision and Recall of the Experiments

6. Conclusions

On the study of songs recommendation algorithms, we need pay more attention to people's inner feelings. For the same user, the feeling about music is closely related to the mood and environment at that time, which user may have a great difference in the feeling of the same song and the feeling changes all the time. In this study, we analyzed the emotion of songs, query songs with similar themes in the music data sets and form a recommendation list based on their listening behavior. In the future we can continue to study the various recommendations, such as the combination of music melody, rhythm, lyricist and other elements. In this way, we will grasp the properties of songs more comprehensively so as to improve the accuracy of recommendation.

Acknowledgements

We thank Baidu Inc. provide lyrics data sets to the open interface, which help us to complete the experiment. This work is also supported by the National Nature Science Foundation of China (Grant No: 61272067), the Natural Science Foundation Team Research Project of Guangdong Province, China (Grant No: S2012030006242).

References

[1] X. Y. Cheng, W. Kang, J. Zhang and Q. Shi, "The research on Chinese word segmentation system with semantic annotations information", The 2nd International Conference on Soft Computing in Information Communication Technology. Atlantis Press, **(2014)**.
[2] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating "word of mouth", Proceedings of the SIGCHI conference on Human factors in computing systems", ACM Press/Addison-Wesley Publishing Company, **(1995)**, pp. 210-217.
[3] S. Jun, D. Kim, M. Jeon, S. Rho and E. Hwang, "Social Mix: An Automatic Music Recommendation and Mixing System based on Social Network Analysis", Journal of Supercomputing, **(2014)**.
[4] H. J. Xu and E. F. Zhang, "Music recommender system based on association rules and singular value decomposition", Electronic Design Engineering, (in Chinese), vol. 21, no. 1, **(2013)**, pp. 17-19.
[5] R. M. Li, H. F. Lin and J. Yan, "Mining latent semantic on User-Tag-Item for personalized music recommendation", Journal of Computer Research and Development, (in Chinese), vol. 51, no. 10, **(2014)**, pp. 2270-2276.
[6] S. S. Liu, "Music recommendation system based on the combination of audio features and social tags", Huazhong University of Science and Technology, (in Chinese), **(2011)**.
[7] J. L. Ouyang, "Graph-based personalized music recommendation", Zhejiang University, (in Chinese), **(2013)**.

Authors

Chengzhou Fu, He graduated in 2012 from School of Software, South China Normal University, with a Bachelor degree, and graduated in 2014 from School of Computer Science, South China Normal University, with a Master degree of Software Engineering, under the supervision of professor Yong Tang. Now he is a PhD student in School of Computer Science, South China Normal University. His research interest is recommendation system.

Yong Tang, He received his BS and MSc degrees from Wuhan University in 1985 and 1990, respectively, and PhD degree from University of Science and Technology of China in 2001, all in computer science. He is a Professor and Dean of the School of Computer Science at South China Normal University. He serves as the Director of Services Computing Engineering Research Center of Guangdong Province. His main research interest is big data.

