Design Engineering

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

A Music Recommendation System based on Reinforcement Learning

Haifei Zhang[1,2], Yiyang Sun[1], Jianlin Qiu[1,3,*], Changyong Zhu[1], Junsong Zhao[1], Haoyu Wang[4]

[1]School of Computer and Information Engineering, Nantong Institute of Technology, Nantong, Jiangsu, China

[2]School of Computer and Information School, Hohai University, Nanjing, Jiangsu, China

[3]School of Information Science and Technology, Nantong University, Nantong, Jiangsu, China

[4]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu, China

*Corresponding Author:Jianlin Qiu

Abstract:

At present, collaborative filtering recommendation method is the most widely used method. However, there are still two key problems in recommender system: cold start, exploration and exploitation. This paper uses multi-armed bandits (ϵ-greedy) algorithm in reinforcement learning to solve the problem of cold start and exploration-exploitation trade-off problem in music recommendation system. Experimental results show that compared with the traditional music recommendation method based on singular value decomposition, the music recommendation system based on this method can better meet the personalized needs of users.

Keywords: *Music recommendation, Reinforcement Learning,Multi-armed bandits,ϵ-greedy, Singular value decomposition.*

I. INTRODUCTION

With the rapid development of Internet technology and big data technology, there are a large number of active users in video websites, information apps, e-commerce websites and so on every day. At the same time, these websites and apps generate a lot of new professional generated content (PGC) or user generated content (UGC) every day (such as novels, information articles, short videos, etc.). For the recommendation system, there are a lot of new and old users, new and old items and a lot of user behavior data. For users, we need to model the users to depict their portraits and interests. However, the behavior of users is sparse, and there may be different proportion of new users. How to recommend to new users is a well-known problem in the recommendation system, namely the cold start problem [1]. What items

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

are displayed to new users determines the user's first feeling and experience. Meanwhile, in the process of recommendation, we need to consider the opportunity to show new items, such as recommending some non-science fiction movies to a user who likes science fiction movies, so as to improve the diversity of recommendation. This is another problem in the recommendation system, namely, the problem of exploration and exploitation [2] trade-off.

The multi-armed bandits algorithm [3] in reinforcement learning [4] is often used to solve the problem of cold start [5] and exploration-exploitation [6] trade-off problem. According to the characteristics of music recommendation, this paper uses multi-armed bandits (ϵ-greedy) methods to solve the problem of cold start and exploration-exploitation [6] trade-off problem in music recommendation system. Compared with the traditional music recommendation method based on singular value decomposition, this method can better meet the personalized needs of users.

II. MULTI-ARMED BANDITS ALGORITHMS

2.1 Thompson Sampling Algorithm

Thompson sampling [7] assumes that the rate of return of each item is p, and beta distribution is used to describe the distribution of return rate. Through continuous tests, a probability distribution based on probability p with high confidence is estimated. Suppose that the probability distribution of probability p conforms to beta (wins, lose). The beta distribution has two parameters, win and lose. Each item maintains the parameters of beta distribution. One item is selected for each test. If there is a click, wins will be increased by 1, otherwise lose will be increased by 1. The way to select an item each time is to use the beta distribution of each item to generate a random number, and select the item with the largest random number among the random numbers generated by all items. The probability density function of beta distribution is shown in formula (1).

$$\mathrm{Beta}(x;\alpha,\beta)=\frac{x^{\alpha-1}*(1-x)^{\beta-1}}{\int_0^1 \mu^{\alpha-1}*(1-\mu)^{\beta-1}du}=\frac{x^{\alpha-1}*(1-x)^{\beta-1}}{B(\alpha,\beta)} \quad (1)$$

2.2 Upper Confidence Bound Algorithm

The steps of the upper confidence bound (UCB) [8] algorithm are as follows:

1). Try all the arms first.

2). Then select the arm with the largest value in formula (2) each time.

$$\bar{x}_j(t)+\sqrt{\frac{2\ln t}{T_{j,t}}} \quad (2)$$

Where t is the current number of tests, $\bar{x}_j(t)$is the average return of this arm from the beginning to the present, $T_{j,t}$is the number of times the arm has been tested, $\sqrt{\frac{2\ln t}{T_{j,t}}}$is called bonus, which is essentially the standard deviation of the mean.

3. Observe the Selection Results and Updatet and $T_{j,t}$.

Formula (2) reflects that the larger the mean value and the smaller the standard deviation,

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

the greater the probability of being selected, which plays the role of exploration; at the same time, those arms that have been selected less times will also get the test opportunity and play the role of exploration.

2.3 Epsilon-Greedy Algorithm

The ϵ-greedy algorithm [9] is based on a probability to make a compromise between exploration and utilization: each attempt, the probability of ϵ is used to explore, that is, a rocker arm is randomly selected with uniform probability; the probability of 1-ϵ is used to exploit, that is, the rocker arm with the highest average reward is selected.

If the uncertainty of rocker arm reward is large, for example, when the probability distribution is wide, more exploration is needed, and a larger ϵ value is needed; if the uncertainty of rocker arm is small, for example, when the probability distribution is concentrated, a small number of attempts can well approximate the real reward, and the required ϵ is small. ϵ is usually given a smaller constant, such as 0.1 or 0.01. However, if the number of attempts is very large, the reward of the rocker arm can be well approximated after a period of time, and there is no need to explore. In this case, ϵ can be gradually reduced with the increase of the number of attempts, $\epsilon = 1/\sqrt{t}$.

III. SINGULAR VALUE DECOMPOSITIONALGORITHM

Singular Value Decomposition (SVD) [10] can be understood as: a more complex matrix A is represented by the multiplication of three smaller and simpler sub matrices, which describe the important characteristics of a large matrix. As shown in formula (3).

$$A = U\sum V^T \tag{3}$$

Where U and V are unit orthogonal matrices, i.e.$UU^T = I$and $VV^T = I$, U is called left singular matrix, V is called right singular matrix, ∑has value only on the main diagonal, we call it singular value, other elements are 0. The dimensions of the above matrices are $U \in R_{m\times n}$, $\sum \in R_{m\times n}$, $V \in R_{m\times n}$.

Generally, ∑has the following forms:

$$\sum = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \end{bmatrix}_{m\times n}$$

When SVD algorithm is applied to recommendation system, the row represents user, column represents item, and the value represents user's rating of item. The advantage of this method is that the user's rating data is sparse matrix. SVD can be used to map the original data into low dimensional space, and then calculate the similarity between items, which can save computing resources.

The overall idea of the recommendation system based on SVD algorithm: first find the items that the user has not scored, and then calculate the similarity between the items not scored and other items in the low dimensional space after SVD "compression", and get a prediction

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

score. Then, sort the scores of these items from high to low, and return the top n items to recommend to users.

IV. EXPERIMENTALRESULTS AND ANALYSIS

4.1 Data Processing

For music datasets, Million Song Dataset (MSD) is well known Data set, which contains more than 1 million songs information, the huge amount of data is about 280gb data files, for such a large data set, identification and acquisition is a more complex process, so this experiment needs to design and compress the data set, which can be more convenient for data storage.

The crawler obtains the data set matrix as shown in Fig 1. The horizontal axis is the type and age information of music, and the vertical axis is the list of all songs. If a certain type or era is satisfied, label 1; if not, label it as 0.

Title	(1980s)	(1990s)	(2000s)	(2010s)	(2020s)	Pop	Rock	Counrty	Folk	Dance	Grunge	Love	Metal	Classic	Funk	Electric	Acoustic	Indie	Jazz	SoundTra	Rap
The Scie	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Yellow	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stairway	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Shape of	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
Fix You	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Wish You	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Smells L	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Hello	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Can't Fe	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comforta	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hymn for	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Paradise	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Adventure	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Viva la V	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cheap Thr	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
In the En	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Sweet Ch	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Californi	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Radioact	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Numb	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Do I Wan	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Creep	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hotline B	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lean On	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Intro	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Hotel Ca	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
A Sky Ful	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Nothing I	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Clocks	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Seven Nat	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig1:Music type data matrix

There are 50 million pieces of data in the MSD data set, which is relatively large. After cleaning the duplicate data and merging the associated data, local sampling is carried out to obtain one million pieces of data (as shown in Fig 2). The processed data is shown in Fig 3.

Design Engineering

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

	track_id	title	song_id	release	artist_id	artist_mbid	artist_name	duration	artis
0	TRMMMYQ128F932D901	Silent Night	SOQMMHC12AB0180CB8	Monster Ballads X-Mas	ARYZTJS1187B98C555	357ff05d-848a-44cf-b608-cb34b5701ae5	Faster Pussy cat	252.05506	
1	TRMMMKD128F425225D	Tanssi vaan	SOVFVAK12A8C1350D9	Karkuteillä	ARMVN3U1187FB3A1EB	8d7ef530-a6fd-4f8f-b2e2-74aec765e0f9	Karkkiautomaatti	156.55138	
2	TRMMMRX128F93187D9	No One Could Ever	SOGTUKN12AB017F4F1	Butter	ARGEKB01187FB50750	3d403d44-36ce-465c-ad43-ae877e65adc4	Hudson Mohawke	138.97098	
3	TRMMMCH128F425532C	Si Vos Querès	SOBNYVR12A8C13558C	De Culo	ARNWYLR1187B9B2F9C	12be7648-7094-495f-90e6-df4189d68615	Yerba Brava	145.05751	
4	TRMMMWA128F426B589	Tangle Of Aspens	SOHSBXH12A8C13B0DF	Rene Ablaze Presents Winter Sessions	AREQDTE1269FB37231		Der Mystic	514.29832	
...	...	...	...	...	...	...	...	...	
999995	TRYYYUS12903CD2DF0	O Samba Da Vida	SOTXAME12AB018F136	Pacha V.I.P.	AR7Z4J81187FB3FC59	9d50cb20-7e42-45cc-b0dd-154c3e92a577	Kiko Navarro	217.44281	
999996	TRYYYJO128F426DA37	Jago Chhadeo	SOXQYIQ12A8C137FBB	Naale Baba Lassi Pee Gya	ART5FZD1187B9A7FCF	2357c400-9109-42b6-b3fe-9e2d9f8e3872	Kuldeep Manak	244.16608	
999997	TRYYYMG128F4260ECA	Novemba	SOHODZI12A8C137BB3	Dub_Connected: electronic music	ARZ3R6M1187B9AF750	8b97e9c8-61f5-4615-9a96-276f24204e34	Gabriel Le Mar	553.03791	
999998	TRYYYDJ128F9310A21	Faraday	SOLXGOR12A81C21EB7	The Trance Collection Vol. 2	ARCMCOK1187B9B1073	4ac5f3de-c5ad-475e-ad50-41f1ef9dba20	Elude	484.51873	
999999	TRYYYVU12903CD01E3	Fernweh feat. Sektion Kuchikäschtli	SOWXJXQ12AB0189F43	So Oder So	AR7PLM21187B990D08	3af2b07e-c91c-4160-9bda-f0b9e3144ed3	Texta	295.07873	

1000000 rows × 14 columns

Fig2: Data set format before processing

	user	song	listen_count	title	song_id	release	artist_name	year
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1	The Cove	SOAKIMP12A8C130995	Thicker Than Water	Jack Johnson	0
1	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAPDEY12A81C210A9	1	Nothing from Nothing	SOAPDEY12A81C210A9	To Die For	Billy Preston	1974
2	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBBMDR12A8C13253B	2	Entre Dos Aguas	SOBBMDR12A8C13253B	Flamenco Para Niños	Paco De Lucia	1976
3	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBFNSP12AF72A0E22	1	Under Cold Blue Stars	SOBFNSP12AF72A0E22	Under Cold Blue Stars	Josh Rouse	2002
4	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBSUJE12A6D4F8CF5	2	12 segundos de oscuridad	SOBSUJE12A6D4F8CF5	10 + Downloaded	Jorge Drexler	2006

Fig3:Data set format of processed

In the music recommendation system, there are many factors have a prominent impact on the relevant music data. From the perspective of users, users' interests are closely related to their gender, age and geographical location. From the perspective of music, the same type of song often has some similarities with singers and composers, and different music types have different audience users. Therefore, the most important thing in recommendation system is user behavior. Fig4 shows the information of a user listening to music.

	user	song	listen_count	title	release	artist_name	year
5	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBXHDL12A81C204C0	1	Stronger	Graduation	Kanye West	2007
7	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODACBL12A8C13C273	1	Learn To Fly	There Is Nothing Left To Lose	Foo Fighters	1999
14	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOFRQTD12A81C233C0	1	Sehr kosmisch	Musik von Harmonia	Harmonia	0
19	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOIZAZL12A6701C53B	5	I'll Be Missing You (Featuring Faith Evans & 1...	No Way Out	Puff Daddy	0
22	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOKRIMP12A6D4F5DA3	5	I?'m A Steady Rollin? Man	Diggin' Deeper Volume 7	Robert Johnson	0

Fig4: The information of a user listening to music

After reading and analyzing the data, the most popular top 20 songs (as shown in Fig 5), the

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

most recent number of user songs played (as shown in Fig 6) and the distribution data of user playing (as shown in Fig 7) are obtained.

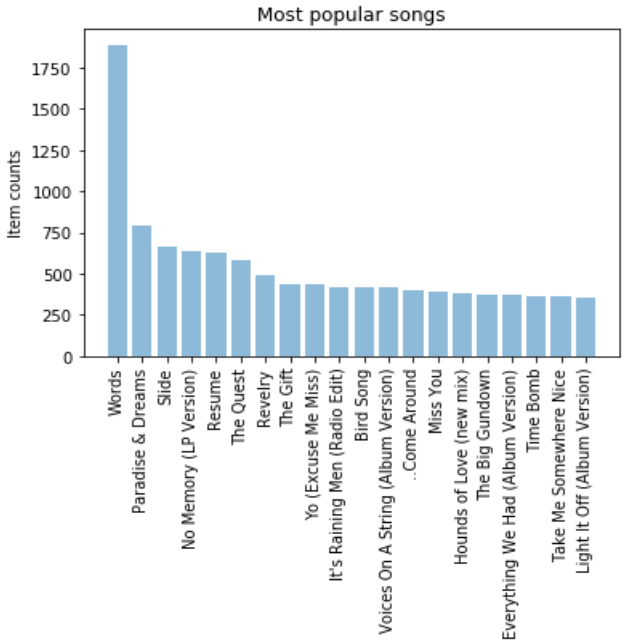

Fig5: The most popular top 20 songs

```
count    9999.000000
mean       49.815882
std        44.772202
min         1.000000
25%        22.000000
50%        37.000000
75%        62.000000
max       695.000000
Name: title, dtype: float64
```

Fig6: The most recent number of user songs played

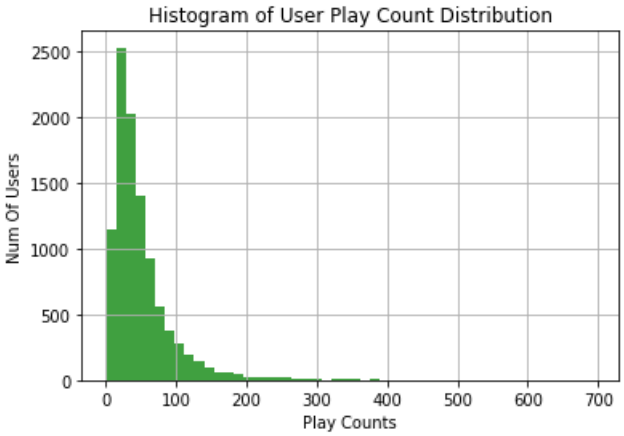

Fig7: The distribution data of user playing

4.2 Music Recommendation based on SVD

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

The experimental data set is given, there is no new song input, so cold start only considers how to recommend to new users. If the basic information of the new user is less, the hot list recommendation is carried out according to the rating from high to low, and then according to the user's preferences. As shown in Fig 8, the first column is the ID value of the data in the dataset, the second column is the name of music, the third column is the cumulative score, and the fourth column is the ranking.

	title	score	Rank
13182	Sehr kosmisch	869	1.0
3851	Dog Days Are Over (Radio Edit)	802	2.0
18293	You're The One	670	3.0
16879	Undo	657	4.0
13161	Secrets	624	5.0
12544	Revelry	620	6.0
6666	Horn Concerto No. 4 in E flat K495: II. Romanc...	553	7.0
6472	Hey_ Soul Sister	515	8.0
5067	Fireflies	512	9.0
16388	Tive Sim	481	10.0
16961	Use Somebody	466	11.0
11071	OMG	426	12.0
2437	Canada	425	13.0
4150	Drop The World	419	14.0
15807	The Scientist	400	15.0
2215	Bulletproof	399	16.0
2817	Clocks	396	17.0
9687	Marry Me	393	18.0
8167	Just Dance	393	19.0
2561	Catch You Baby (Steve Pitron & Max Sanna Radio...	384	20.0

Fig8: Cold start recommendation list

After solving the cold start problem, music recommendation is conducted again. There are nearly 50 million pieces of data in the data set. Although one million pieces of data are selected for the experiment, the loading process is still slow. Therefore, 500 pieces of data are selected for sampling and recommendation. The data of a user's listening to music is shown in Fig 4. Since there is no rating data, the number of times a user listens is taken as the score value. According to the user's click times, the music that you like is the most.

According to the data read as shown in Fig 4, the number of times of listening (listen_ count) is converted to score, and the matrix of all users, songs and score values is established. The SVD algorithm is used to recommend songs to users, as shown in Fig 9.

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

```
Recommendation for user with user id 5
The number 1 recommended song is Halo BY The Pussycat Dolls
The number 2 recommended song is The Baby Screams BY The Cure
The number 3 recommended song is Hallelujah (Album Version) BY Paramore
The number 4 recommended song is Seven Nation Army BY The White Stripes
The number 5 recommended song is Hallelujah (Album Version) BY Paramore
The number 6 recommended song is Slip Away BY Clarence Carter
The number 7 recommended song is Mass Appeal (Explicit) BY Gang Starr
The number 8 recommended song is The Perfect Kiss BY New Order
The number 9 recommended song is A Message To You Rudy BY The Specials
The number 10 recommended song is This Is A Forgery BY Dashboard Confessional
```

Fig9: Music recommendation results based on SVD algorithm

4.3 Music Recommendation Based on $\varepsilon -$ GreedyAlgorithm

In the MSD dataset containing more than 1 million songs, select the music type you like (multiple choices are allowed), as shown in Fig 10.

```
Select song features that you like
1. Rock
2. Country
3. Folk
4. Dance
5. Grunge
6. Love
7. Metal
8. Classic
9. Funk
10. Electric
11. Acoustic
12. Indie
13. Jazz
14. SoundTrack
15. Rap
Enter number associated with feature:
```

Fig10: Select music type

According to the selected music type, the system first publishes several songs for audition and scoring (as shown in Fig 11), and then makes music recommendation according to the user's score, as shown in Fig 12.

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

```
Enter number associated with feature: 1
Do you want to add another feature? (y/n) y
Enter number associated with feature: 13
Do you want to add another feature? (y/n) n


Rate following 5 songs. So that we can know your taste.

How much do you like "Ramble On" (1-10): 1
How much do you like "In My Place" (1-10): 5
How much do you like "White Winter Hymnal" (1-10): 7
How much do you like "Shape of You" (1-10): 2
How much do you like "Mykonos" (1-10): 10
```

Fig11: Score the released music

```
Wait


1. Heathens
2. Dirty Paws
3. Complicated
4. Stubborn Love
5. Why'd You Only Call Me When You're High?
6. Clocks
7. The Sound of Silence
8. Holocene
9. Like a Rolling Stone
10. Ho Hey
```

Fig12: Music recommendation resultsbased on ϵ-greedy algorithm

If the user is satisfied with the recommendation result, he can choose to quit the system; if the user is not satisfied with the recommendation result, he can choose to recommend again, and the system will automatically launch songs for users to listen to and score. As shown in Fig 13.

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

```
Rate songs one by one or leave it blank
Rate Take Me Out (1/10): 1
Rate Somewhere Only We Know (1/10): 3
Rate Drive (1/10): 7
Rate Way Down We Go (1/10): 5
Rate Why'd You Only Call Me When You're High? (1/10): 4
Rate Time (1/10): 1
Rate New Slang (1/10): 2
Rate It's Time (1/10): 3
Rate Tighten Up (1/10): 7
Rate Fluorescent Adolescent (1/10): 5

Do you want more recommendations? (y/n)
```

Fig13: Recommendation once again

4.4 Experimental Analysis

The music recommendation system based on SVD algorithm needs to read and write a large amount of data. Through the visual interface, we can see the change track of the data and the number of clicks on music. In the case of cold start, recommendation directly according to the list does not necessarily meet the needs of users. According to the number of hits, there may be careless clicks by users. However, the music recommendation system based on the multi-armed bandit ($\varepsilon -$ Greedy) algorithm describes the interactive recommendation problem as a context multi-armed bandits. The system recommends new songs according to users' preferences and obtains their personalized scores to reduce cold start, which is more objective.

V. CONCLUSION

In this paper, the music recommendation systems is implemented by using multi-armed bandits (ϵ-greedy) method in reinforcement learning and singular value decomposition algorithm in traditional machine learning. The SVD algorithm uses a small data set to represent the original data set, which can remove noise and redundant information, so as to achieve the purpose of optimizing the data. For the cold start problem, the system can only recommend through the ranking list information which is generated according to the number of hits, and cannot meet the individual needs of users, and cannot solve the problem of exploration and exploitation. The music recommendation method based on the multi-armed bandits (ϵ-greedy) algorithm describes the interactive music recommendation problem between users and the system as the multi-armed bandits problem. According to users' preferences, the system recommends new songs and obtains their scores, which can better alleviate problem of cold start and exploration-exploitation trade-off, and can better meet the personalized needs of users.

ACKNOWLEDGEMENTS

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

This work is sponsored by: (1) the team for science & technology innovation and local development service of Nantong Institute of Technology under Grant No.KJCXTD312; (2) the Universities Natural Science Research Project of Jiangsu Province under Grant No.17KJB520031; (3) the science and technology planning project of Nantong City under Grants No.YYZ17078 and No.MSZ18030; (4) the scientific research backbone training project of Nantong Institute of Technology under Grant No. ZQNGG109; (5) Business Administration first class built subject of Jiangsu province in 13th Five-Year under Grant No. SJY201609; (6)logistics management brand specialty project of Nantong Institute of Technology under Grant No.NTLG201624; (7) the Universities Humanities and Social Sciences outside school research base—Nantong Shanghai Industrial collaborative development research base under Grant No. SJSZ201716.

REFERENCES

[1] Ke Z, Yang S. H., Zha H (2011) Functional Matrix Factorizations for Cold-Start Recommendation. International AcmSigir Conference on Research & Development in Information Retrieval ACM: 315–324, doi: 10.1145/2009916.2009961

[2] Agarwal D, Chen B, Elango P (2009) Explore/Exploit Schemes for Web Content Optimization. 2009 Ninth IEEE International Conference on Data Mining, Miami, FL: 1-10,doi: 10.1109/ICDM.2009.52

[3] Joannès V., Mohri M (2005) Multi-armed Bandit Algorithms and Empirical Evaluation.European Conference on Machine Learning, Springer-Verlag: 437-448

[4] Sutton R, Barto A (2018) Reinforcement Learning: An Introduction, Second Edition. Cambridge,MA: MIT Press.ISBN978-0-262-19398-6

[5] Felicio C., Paixão K., Barcelos C., Preux P. (2017) A Multi-Armed Bandit Model Selection for Cold-Start User Recommendation.In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP ’17), Association for Computing Machinery, New York, NY, USA: 32-40, doi: 10.1145/3079628.3079681

[6] Avner O., Mannor S., Shamir O. (2012) Decoupling Exploration and Exploitation in Multi-Armed Bandits. Proceedings of the 29th International Conference on Machine Learning(ICML’12), Omnipress, Madison, WI, USA: 1107-1114.

[7] Timothy V., Eugenio B., Pieter J K L (2020) Thompson Sampling for Factored Multi-Agent Bandits. Proceedings of the 19th International Conference on Autonomous Agents and Multi Agent Systems (AAMAS’20), Auckland New Zealand: 2029-2031

[8] Huang K.H., Lin H.T. (2016) Linear Upper Confidence Bound Algorithm for Contextual Bandit Problem with Piled Rewards.PAKDD 2016: Proceedings, Part II, of the 20th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining-Volume 9652:143-155, doi: 10.1007/978-3-319-31750-2_12

ISSN: 0011-9342
Issue: 11 | Pages: 331 - 342

[9] Sammut C, Webb G.I.eds (2017) Multi-armed Bandit. In: Encyclopedia of Machine Learning and Data Mining. Boston, MA: Springer. ISBN978-1-4899-7685-7, doi: 10.1007/978-1-4899-7687-1_100315

[10] Aggarwal C.C (2020) Singular Value Decomposition. In: Linear Algebra and Optimization for Machine Learning. Cham, Swiss: Springer. ISBN978-3-030-40343-0, doi:10.1007/978-3-030-40344-7_7