



INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

RECOMMENDATION SYSTEM WITH DETECTION AND PREVENTION OF MALICIOUS FEEDBACK RATING

Mr. Satyadeep Shamrao Basugade *, Mr. Vipul Vinayak Bag

* Department of Computer Engineering, N.K.Orchid College of Engineering, Maharashtra, India

ABSTRACT

Recommendation system can help users to provide right items from large number of available items. Recommender systems suggest items to users by using the techniques of Collaborative filtering based on historical data of items that users have rated. In this paper we present a novel collaborative filtering approach called NPC collaborative filtering for item i.e books recommendations with malicious feedback rating detection and prevention system. Goal of detection and prevention system is to detect the malicious feedback rating and avoid or adjust this malicious feedback rating. The experimental results show that our approach achieves better accuracy than other competing approaches.

KEYWORDS: Collaborative filtering, similarity measures, Recommender system, Feedback rating, Exponentially Weighted Moving Average.

INTRODUCTION

Recommender system can help user to solve the problem by providing them with personalized suggestions. There are different types of recommendation systems including content-based, collaborative and hybrid recommendation [2]. Collaborative filtering (CF) approaches are widely adopted for the recommender systems [1]. Collaborative filtering algorithms can be grouped into two classes: memory-based and model-based. Memory-based collaborative filtering approaches are usually classified into user-based approaches item based approach and their combined approaches, Similarity measures have been discussed in several investigations in memory-based collaborative filtering approaches, the Pearson correlation coefficient (PCC) and the Normal Recovery approach (NR) are the two most popular approaches to measure the similarity [4]. It is difficult to ensure the true value of user feedback ratings because of the presence of malicious users. Malicious users could provide malicious feedback ratings to affect the measurement results for commercial benefit. The PCC does not properly handle the difference of vectors in different vector spaces [1]. In existing system [1] it does not deal with malicious feedback rating. Previous approaches fail to ensure the accuracy of feedback ratings [3]. Users have different feedback rating styles. Different users often give different feedback ratings to the same item. For a reputation mechanism to be fair and objective, it is essential to measure reputation on the basis of fair and objective feedback ratings.

MATERIALS AND METHODS

To build recommender system with detection and prevention of malicious feedback rating, we have design new item-similarity and user-similarity using similarity measure. Then create clusters of that similar items and users using k-nearest neighbour clustering methods. Then NPCF recommendation is used to generate recommendation and then given to users. After that users will give feedback rating. Then detect the malicious feedback rating. Then feedback similarity computation is performed. After detecting malicious user prevent user means block that user. Then get the adjusted rating and update the database. On the basis of that updated data recommendation is generated.

Compared with the previous work, we propose a new similarity measurement approach and a novel collaborative filtering approach, named NPCCF. The contributions of this paper can be summarized as follows:

- We design a new user and item based similarity measure for memory-based collaborative filtering.
- We propose a new collaborative filtering approach, which significantly improves the recommendation performance compared with the other well-known approaches.
- We design a new detection and prevention system for malicious users, which will help to improve recommendation performance.

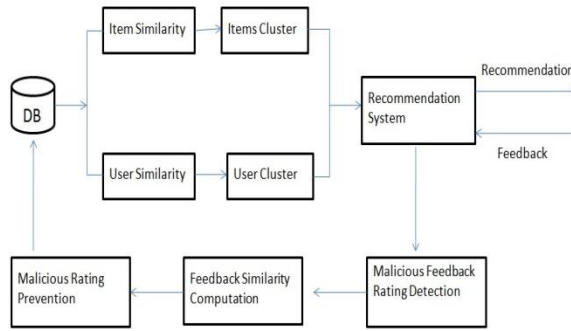


Fig 1 System Overview

PROPOSED MODELLING

In this section we present our new design similarity measures, collaborative filtering, and malicious user detection and prevention approach. Given recommendation system that contains M users and N items (Books), we obtain an $M \times N$ user-item matrix, in which entry $r_{m,n}$ denotes the rating of item n given by user m. If entry $r_{m,n}$ is empty, then $r_{m,n} = \emptyset$, denoting that the item n has never been rated by user m before.

NPC Similarity Measures:

As shown in Fig. 2, Let us consider a user-item matrix. In this there are 5 users (u_1 to u_5) and 5 items (i_1 to i_5) where 5 and 1 are highest and lowest rating respectively.

User / Items	i_1	i_2	i_3	i_4	i_5
u_1	1	2	3	4	5
u_2	2	2	3	4	
u_3	3	2		4	
u_4	1	1	1	1	1
u_5	5	5	5	5	5

Fig. 2 Motivating Example

After computing similarity we get result, u_1 is more similar to u_2 than u_3 . But as shown in Fig.2, u_1 is actually less similar to u_2 than u_3 . Because the rating of u_2 and u_3 both are between 2 and 4 and the rating of u_1 is between 1 and 5. So from that we could say that PCC does not properly handle the rating difference between users.

By Applying NR similarity measure we get result, for computing similarity between u_4 and u_5 the equation (2) fail to work. Because if $r_{u_{max}} = r_{u_{min}}$ then denominator goes to 0 and we cannot divide by 0. In matrix user u_4 having $r_{u_{max}} = 1$ and $r_{u_{min}} = 1$, so as for user u_5 having $r_{v_{max}} = 5$ and $r_{v_{min}} = 5$. To overcome these drawbacks, we propose new similarity measure approach i.e NPC. In our approach to measure the similarity between users, we first normalize each row of the original user-item matrix T by the highest and lowest rating of the same row, So that each row vary between 0 to 1.

User / Items	i_1	i_2	i_3	i_4	i_5
u_1	0	0.25	0.50	0.75	1
u_2	0	0	0.50	1	
u_3	0.50	0		1	
u_4	∞	∞	∞	∞	∞
u_5	∞	∞	∞	∞	∞

Fig. 3 Normalize Matrix for User

The NPC can be employed to measure similarity between two users u and v by,

$$Sim(u, v) = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \tag{1}$$

$$\frac{\sum_{i \in I} \left(\frac{r_{u,i} - r_{u_{min}}}{r_{u_{max}} - r_{u_{min}}} - \frac{r_{v,i} - r_{v_{min}}}{r_{v_{max}} - r_{v_{min}}} \right)^2}{\sqrt{|I|}}$$

Where $I = I_u \cap I_v$ is the set of items rated by users u and v , $r_{u,i}$ is the rating of item i given by user u , \bar{r}_u denotes average rating of user u on items in I . Where $I = I_u \cap I_v$ is the set of items rated by users u and v . $|I|$ is the number of items, $r_{u,i}$ is the rating of item i from user u . $r_{u_{max}}$ and $r_{u_{min}}$ are the highest and lowest rating given by user u respectively. $r_{v_{max}}$ and $r_{v_{min}}$ are the highest and lowest rating given by user v respectively. The PCC value ranges between -1 to 1. The NR value ranges between 0 to 1. So we map the -1 to 1 range into 0 to 1 range. So the NPC value ranges between 0 to 1. If $Sim(u, v) = 0$ then two users are dissimilar. If $Sim(u, v) = 1$ then two users are similar.

In our approach to measure the similarity between items, we first normalize each column of the original user-item matrix T by the highest and lowest rating of the same column, So that each column vary between 0 to 1.

User / Items	i_1	i_2	i_3	i_4	i_5
u_1	0	0.25	0.50	0.75	1
u_2	0.25	0.25	0.50	0.75	
u_3	0.50	0.25		0.75	
u_4	0	0	0	0	0
u_5	1	1	1	1	1

Fig. 4 Normalize Matrix for Item

The NPC can be employed to measure similarity between two items i and j by,

$$Sim(i, j) = \left\{ \begin{array}{l} \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \\ 1 - \sqrt{\frac{\sum_{u \in U} (\frac{r_{u,i} - r_{imin}}{r_{imax} - r_{imin}} - \frac{r_{u,j} - r_{jmin}}{r_{jmax} - r_{jmin}})^2}{|U|}} \end{array} \right. \quad (2)$$

Where $U = U_i \cap U_j$ is the group of users who rated both items i and j , $r_{u,i}$ is the rating of item i given by user u , \bar{r}_u denotes average rating of user u on items in I . Where $U = U_i \cap U_j$ is the set of users who rated both items i and j . $|U|$ is the number of users, $r_{u,i}$ is the rating of item i from user u . r_{imax} and r_{imin} are the highest and lowest rating of item i respectively. r_{jmax} and r_{jmin} are the highest and lowest rating of item j respectively. The PCC value ranges between -1 to 1. The NR value ranges between 0 to 1. So we map the -1 to 1 range into 0 to 1 range. So the NPC value ranges between 0 to 1. If $Sim(i, j) = 0$ then two items are dissimilar. If $Sim(i, j) = 1$ then two items are similar. When PCC fails we use the NR approach and when NR fails we use difference technique as explain below.

As shown in Fig.3 for user u_4 and u_5 having normalize value as ∞ . It means that their rating of $r_{imax} = r_{imin}$.

For these kinds of situations we have decided that to take the differences of their rating. When the difference is less then the similarity is greater i.e users are similar. If difference is more the similarity is less i.e users are dissimilar. After getting similar user and item we form a cluster i.e similar user cluster and similar item cluster using k-nearest neighbor clustering method

.NPC Collaborative Filtering:

Based on our NPC similarity measurement approach, we propose an innovative collaborative filtering method, NPC collaborative filtering (NPCCF). For predicting the unknown rating $\hat{r}_{u,i}$ of user u on item i , we propose our user-based NPCCF as follows:

$$\hat{r}_{u,i} = r_{umin} + (r_{imax} - r_{umin}) \frac{\sum_{u' \in U} Sim(u, u') \times nr_{u',i}}{\sum_{u' \in U} Sim(u, u')} \quad (3)$$

Here $Sim(u, u')$ can be computed by (1) where U are set of similar users to user u , who have rated item i , $nr_{u',i}$ is the rating of item i from user u' .

In item-based rating prediction, we define item-based NPCCF formula as

$$\hat{r}_{u,i} = r_{imin} + (r_{imax} - r_{imin}) \frac{\sum_{i' \in I} Sim(i, i') \times nr_{u,i'}}{\sum_{i' \in I} Sim(i, i')} \quad (4)$$

Here $Sim(i, i')$ can be computed by (2) where I are set of similar items to item i , who have rated by user u , $nr_{u,i'}$ is the rating of item i' from user u .

To use both user-based and item-based prediction at same time, we create a combination of these formulas as below:

$$\hat{r}_{u,i} = r_{umin} + \lambda \times (r_{imax} - r_{umin}) \frac{\sum_{u' \in U} Sim(u, u') \times nr_{u',i}}{\sum_{u' \in U} Sim(u, u')} + (1-\lambda) \times (r_{imax} - r_{imin}) \frac{\sum_{i' \in I} Sim(i, i') \times nr_{u,i'}}{\sum_{i' \in I} Sim(i, i')} \quad (5)$$

Malicious Rating Detection:

This section focuses on the application of The Exponentially Weighted Moving Average (EWMA) to

detect and handle the malicious feedback rating as follows:

$$EWMA_t = \lambda Y_t + (1 - \lambda) EWMA_{t-1} \text{ for } t=1,2,3,\dots,n \quad (6)$$

Where $EWMA_0$ is the target rating. It is calculated by taking average of all rating given by user u to items i . Y_t is rating given by user u . n is number of items to be monitored including $EWMA_0$. λ is vary between 0 to 1.

The estimated variance of the EWMA is

$$s_{ewma}^2 = \frac{\lambda}{2-\lambda} s^2 \quad (7)$$

Where s is the standard deviation calculated from given data. t is not small. The center line is the target value or $EWMA_0$. The lower and upper limites are:

$$UCL = EWMA_0 + k s_{ewma} \quad (8)$$

$$LCL = EWMA_0 - k s_{ewma} \quad (9)$$

Where λ is set as 1 and factor k is set as 3. First using equation (10) we have to calculate the actual plotting point according to given data. Then we have to calculate upper and lower limits of ratings. if the plotted points are lies between the UCL and LCL the they are consider as true rating. if they are beyond the UCL and below the LCL then they are consider as malicious or false or fake ratings. from that we can detect the malicious ratings.

Malicious Rating Prevention:

In this section, we prevent malicious feedback rating for that we propose malicious feedback rating prevention scheme.

To gain true rating we must do rating adjustment. In this the users which are partially malicious are considered and their malicious rating is to be adjusted. For this we have to get the similar users of that malicious user that we can get by computing equation (5). Now we have to pick up the similar users cluster from user cluster. To adjust feedback rating of user a according to the feedback rating of other similar users with the following:

$$\hat{r}_{a,i} = \sum_{u \in S(a)} \frac{S^k(a,u)}{\sum_{u \in S(a)} S^k(a,u)} \times r_{u,i} \quad (10)$$

Where S^k is the set of similar users. $S^k = \{ S_1^k, S_2^k, \dots, S_l^k \}$ which contains l items used by the K users. $S^k(a,u)$ is the similarity value of user a and user u . $\hat{r}_{a,i}$ is the adjusted feedback rating of i -th rated item from user a , $r_{u,i}$ is the rating of item i rated by user u .

In this prevention scheme we does not consider the all rating given by total malicious user. These rating are totally discarded from system so that we can get true value of that particular items.

RESULTS AND DISCUSSION

Experiment Setup:

The experiment is conducted on book dataset. For this experiment we use the LensKits books dataset. In this dataset they provided the users, rating, isbn number, book title, year of publication, image url, book author.

Evaluation Metric:

To evaluate the rating prediction accuracy, we use the mean absolute error (MAE). The MAE is the average absolute deviation of predictions to the ground true rating. The MAE is define as

$$MAE = \frac{\sum_{u,i} |r_{u,i} - \hat{r}_{u,i}|}{N} \quad (11)$$

Where $\hat{r}_{u,i}$ denotes the predicted rating of item i for user u . N denotes the total numbers of predicted ratings. $r_{u,i}$ Denotes the actual or true rating of item i rated by user u . The Smaller values indicate better prediction accuracy.

Performance Comparison of Similarity Measures:

To show performance effectiveness of our NPC similarity measures, we compare it with two other similarity measures i.e PCC and NR. We combined PCC, NR and NPC with the formula shown in (5) for missing rating prediction. In this experiment we vary value of λ from 0.1 to 1 with step value of 0.1 Table 1 shows the prediction accuracy of different similarity measures. From Table 1 we can see that the best MAE of PCC, NR and NPC are 3.6303, 3.4946, and 3.4314

respectively. Compared with other approaches our approach significantly improves the prediction accuracy. The Fig. 5 shows MAE performance. Where X-axis shows λ values and Y-axis shows MAE values.

Table 1. MAE Performance Comparison of Different Similarity Measures (Smaller Value Means Better Performance)

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
PCC	4.328	4.2384	4.1595	4.0838	4.008	3.9322	3.8565	3.7807	3.7049	3.6303
NR	4.2837	4.1919	4.1	4.0082	3.9164	3.8245	3.7327	3.6421	3.5677	3.4946
NPC	4.2933	4.1974	4.1014	4.0055	3.9096	3.8136	3.7177	3.6217	3.5258	3.4314

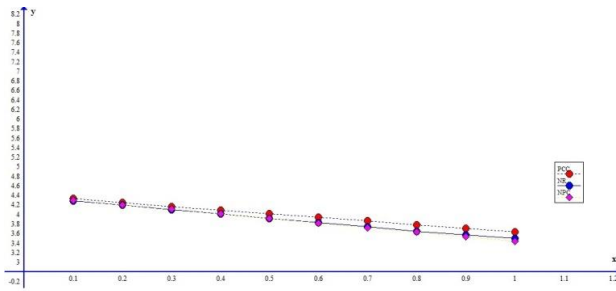


Fig. 5 MAE Performance

Detection of Malicious Users:

There are three types of users

True User:

The user which gives all the rating within the UCL and LCL. Which are true ratings. The Fig. 6 shows True User. In this X-axis shows the number of items rated by that user. Y-axis shows rating. As we see all rating given by user is lies between LCL i.e Lower Control Limit and UCL i.e Upper Control Limit.



Fig. 6 True User

Partially Malicious User:

The partially malicious user and total malicious user are decided on the bases of how many number of malicious rating and number of true rating given by user. The Fig. 7 shows partially malicious user. In this the numbers of true ratings i.e within LCL and UCL are 11 and malicious rating i.e Above UCL or Below LCL are 9. So 11 > 9 it means that this user is partially malicious user. For this kind of users we adjusted the malicious rating and consider it into further calculations.



Fig. 7 Partially Malicious User

Total Malicious User:

The user which gives all or most of his ratings below LCL or Above UCL indicate its Total malicious user. Fig 8 shows Total malicious user. In this user rated only one true rating and rest of ratings are malicious. It means that number of true rating < number of malicious rating. It means that this user is total malicious user. For this kind of user we completely discard all its rating from calculations. It means that block that malicious user. So that, we can gain true rating of items.

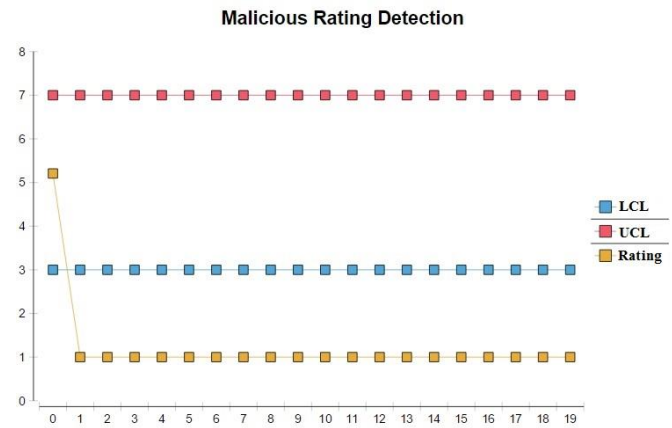


Fig. 8 Total Malicious User

CONCLUSION

In this paper, we propose a NPCCF approach to address problem of item recommendation. This approach finds similar user and items more accurately and leads to better rating prediction accuracy. The proposed malicious feedback rating detection and prevention approach improves the item recommendation process by excluding malicious rating and gaining true rating of that item.

ACKNOWLEDGEMENTS

It was highly eventful at the department of Computer Science and engineering, Nagesh Karjagi Orchid College of Engineering Technology, Solapur. Working with highly devoted professor community will retain a memorable experience. Hence this acknowledgement is humble attempt to honestly thank all those who are directly or indirectly involved in my dissertation work and are of immense help to me. I would sincerely like to thank you my guide Prof. V. V. Bag for giving me perspective and taking interest in this dissertation work and whose advice and teaching helped me adopting a more pragmatic approach.


REFERENCES

- [1] Huifeng Sun, Zibin Zheng, Member, IEEE, Junliang Chen, and Michael R. Lyu, Fellow, IEEE "Personalized Web Service Recommendation via Normal Recovery Collaborative Filtering" IEEE transactions on services computing, vol. 6, no. 4, october-december 2013
- [2] A. C. M. Fong, *Senior Member*, IEEE, Baoyao Zhou, S. C. Hui, *Senior Member*, IEEE, Guan Y. Hong, and The Anh Do " Web Content Recommender System based on Consumer Behavior Modeling" IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, May 2011.
- [3] Shanguang Wang, Member, IEEE, Zibin Zheng, Member, IEEE, Zhengping Wu, Member, IEEE, Fangchun Yang, Member, IEEE, Michael R. Lyu, Fellow, IEEE "Reputation Measurement and Malicious Feedback Rating Prevention in Web Service Recommendation Systems" IEEE transactions on services computing, vol. , no. , march 2014
- [4] Mohd Hilmi Hasan, Jafreezal Jaafar, and Mohd Fadzil Hassan Universiti Teknologi PETRONAS, Tronoh 31750 Perak, Malaysia "Fuzzy-based Clustering of Web Services' Quality of Service: A Review " Journal of Communications Vol. 9, No. 1, January 2014.
- [5] Mohammad Yahya H. Al-Shamri, Nagi H. Al-Ashwal "Fuzzy-Weighted Similarity Measures for Memory-Based Collaborative Recommender Systems " Journal of Intelligent Learning Systems

and Applications, 2014, 6, 1-10 Published Online February 2014.

- [6] P. Resnick and H. R. Varian, "Recommender systems," Communications of the ACM, vol. 40, no. 3, pp. 56–58, 2010.
- [7] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 143–177, 2008.
- [8] Francesco Ricci · Lior Rokach · Bracha Shapira · Paul B. Kantor. "Recommendation System handbook" http://www.cs.bme.hu/nagyadat/Recommender_systems_handbook.pdf
- [9] Cai-Nicolas Ziegler, DBIS Freiburg. "Book-Crossing Dataset" <http://www2.informatik.unifreiburg.de/~chiegler/BX/>
- [10] Joseph A Konstan, Michael Ekstrand. "Introduction to recommender system." <https://www.coursera.org/learn/recommender-systems>
- [11] Jnanamurthy HK, Sanjay Singh "Detection and Filtering of collaborative malicious user in reputation system using quality repository approach" arXive :1308.3876v1 [cs.SI] august 2013

AUTHOR BIBLIOGRAPHY

	<p>Mr. Satyadeep Basugade</p> <p>He obtained his bachelor's degree from the Department of technology, Shivaji University, Kolhapur. He is currently a Master of Engineering (M.E.) student under the supervision of Prof. Vipul Bag. His research is centered on recommendation system and detection and prevention of malicious feedback rating.</p>
--	--



Mr. Vipul Bag

He received M.Tech in Computer Science & Technology and pursuing Ph.D. He is associate professor & head of CSE dept. He has 16 years of teaching experience. He has co-authored over 20 International Journal Publications. The current research interests of Professor Bag's include: 1) Recommendation systems, 2) Data Mining, 3) Computer Network, 4) Stock Market Technical Analysis, 5) Classification (Machine Learning), and 6) Time Series. He was ex- member of Adhoc Board of studies (CSE) Solapur Univesity, Solapur.