Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.26, NO.4, Aug. 2016 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2016.26.4.953

이분법 선호도를 고려한 강건한 추천 시스템*

이 재 훈,[1†] 오 하 영,[2] 김 종 권[1‡]
[1]서울대학교, [2]아주대학교

Bipartite Preference aware Robust Recommendation System*

Jaehoon Lee,[1†] Hayoung Oh,[2] Chong-kwon Kim[1‡]
[1]Seoul National University, [2]Ajou University

요 약

온라인 시스템이 활성화 되고 접근 가능한 정보의 양이 늘어나면서 추천 시스템의 영향력 또한 커지고 있다. 하지만 일부 악의적인 유저들의 공격으로 인해 시스템에 대한 신뢰도를 저하시키고 조작하려는 시도가 늘고 있다. 본 연구팀은 해당 리뷰에 대한 공감, 비공감 비율을 분석하고 이를 추천 시스템에 적용함으로써 추천 시스템의 성능을 향상시키고 강건한 시스템을 유지할 수 있는 방법을 제안한다. 실제 영화 데이터를 수집하여 적용해 본 결과 기존의 추천 시스템보다 향상된 성능을 보였다.

ABSTRACT

Due to the prevalent use of online systems and the increasing amount of accessible information, the influence of recommender systems is growing bigger than ever. However, there are several attempts by malicious users who try to compromise or manipulate the reliability of recommender systems with cyber-attacks. By analyzing the ratio of 'sympathy' against 'apathy' responses about a concerned review and reflecting the results in a recommendation system, we could present a way to improve the performance of a recommender system and maintain a robust system. After collecting and applying actual movie review data, we found that our proposed recommender system showed an improved performance compared to the existing recommendation systems.

Keywords: Recommendation system, Sybil attack, Movie site crawling

I. Introduction

Due to the recent prevalence of social network services such as Facebook and Twitter, the amount of information accessible for us is growing bigger than ever. Studies have been constantly carried out on recommendation system algorithms to provide the list of carefully-selected items that consumers want the most amid the deluge of information, and some of them are actually used by Watcha and Netflix. Through a recommendation system, a business can boost consumer's motivation to purchase a given item, and can reduce consumer's troubles to decide what to buy by proposing them an easy

Received(01. 05. 2016), Modified(06. 28. 2016), Accepted(08. 09. 2016)

* 본 연구는 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단 (No. NRF - 2014R1A1A1003562), 2015년도 정부(미래창조과학부)의 재원으로 한국 연구재단 (No. NRF- 2015R1A2A1A01007400), 2016년도 정부(미래창조과학부)의 재원으로 한국 연구재단 (No.2016R1A5A1012966)의 지원을 받아 수행된 기초연구사업임

† 주저자, jhlee@popeye.snu.ac.kr
‡ 교신저자, ckim@snu.ac.kr(Corresponding author)

access to the items that they want.

However, as recommender systems require a simple subscription procedure and allow users to post their reviews about items without additional security procedures, some malicious users can intentionally give unreasonably higher ratings to a newly released movie than it actually deserves in order to attract the attention of movie goers, or give low ratings to belittle the concerned movie to or discourage consumers from watching it. This is called 'Sybil Attack'[1], and if a website is under a Sybil attack, the reliability of the recommender system of the concerned one will be severely compromised due to the influence of the distorted information and cannot accurately provide a list of recommended items, and the deteriorating reliability of the recommendation system, consumers will not trust the recommended items. In recent years, some robust recommendation algorithms to defend against Sybil attacks have been proposed[2-7].

The Sybil attacks can be mainly divided into push attacks, which give maximum ratings to the concerned item, and nuke attacks, which give minimum ratings. Sybil attackers often use various techniques to hide their identities, such as 'Random Attack', 'Average attack' and 'Bandwagon Attack' and to reduce the influence of the Sybil defense mechanism. Also, it is a very challenging task to distinguish Sybil attacks from the unusual tendency of users who give the extremely opposing ratings.

Our proposed recommender system allows users to check the average rating of a concerned movie in addition to the bipartite reviews (sympathy or apathy) about the concerned rating. Naver movie webpage allows users to post their ratings about an item and write their comments about the concerned item to provide the foundation for their own ratings. In addition, the website can secure the reliability of its recommendation system by allowing users to evaluate other user's ratings with sympathy or apathy responses.

The website can defend itself from Sybil attacks by evaluating the reliability of a concerned movie review with the ratio of sympathy responses against apathy responses. No matter how hard Sybils try to manipulate the average rating about a movie review, it will be impossible unless it earns the sympathy of other users.

Therefore, we first analyzed the patterns of Sybil attacks with the bipartite preference. We obtained a much better performance when we applied the bipartite preference to the Matrix Factorization (MF) technique, a model-based recommendation algorithm. For this experiment, we directly crawled information from Naver movie which is one of the Korea's famous website.

The major contributions of this paper can be summarized as follows.

1. Different from strong assumption of previous many works, we assume initial crawled ground truth dataset can include Sybils.

2. The reliability of the recommender system about a movie review was secured by analyzing the ratio of sympathy responses against apathy responses.

3. We develop a more accurate recommender system by enabling the recommendation system to thwart the influence of Sybils.

4. First we try to differentiate between Sybils and unusual users with a correlation between additional short answers (i.e., bipartite preferences) and

original rating value per each user and item pair. Since previous work does not try to do this part, they easily delete unusual rating information resulting in reducing the inaccurate recommendation.

The remainder of this paper is organized as follows. In Chapter 2, we reviewed the related studies. In Chapter 3, we explained the proposed recommendation system. In Chapter 4, we described the experiment procedures and results. Lastly in Chapter 5, we provided the conclusions of this paper.

II. Related work

2.1 Robust Recommender System

Ever since questions were first raised about Sybil attacks[2], which forged a countless number of IDs with a malicious intention to disrupt a recommender system, many related studies have been carried out. In [3], the study provided the definitions of various attacks including push attack and nuke attack and conducted an experiment with the actual data based on this. In [4], the study conducted the detection of five types of attacks by malicious users by using the K-NN, K-Means Clustering, and PLSA (Probabilistic Latent Semantic Analysis) techniques. In [5], the detection of abnormal users was carried out using the supervised classification technique which utilized the distribution and similarity. In [6], the study proposed an algorithm for detection of obfuscation attacks. A probabilistic detection technique against the recent Sybil attack patterns such as Random, Average, and Bandwagon attacks, has been proposed in [7].

2.2 Sybil detection scheme

Authors of [9] present a spectral clustering method to make recommender systems resistant to Sybil profiles with high correlation by finding the min-cut solution in graph modeling. The edge density of graph modeling allows dealing with an unbalanced clustering. Based on above procedure, finally they define the problem as finding a maximum sub-matrix in the user-user similarity matrix.

Author of [10] first introduce SybilBelief, a semi-supervised learning framework, to detect Sybil nodes with incorporating and propagating known labels. This assumption is a practical since for example, in Twitter, verified users can be treated as known benign labels and users spreading spam or malware can be treated as known Sybil labels[10]. In addition, they focus on another important part of tolerating label noise concept. Even though the labeling is not correct, the authors designed resilient SybilBelief with the probabilistic approach.

III. System model

3.1 Motivation

We analyzed the data crawled from Naver movie webpage by ourselves to find out what kind of influence the proposed recommender system would have on sympathy and apathy responses, and the results are summarized as seen in Table

Table 1. Sympathy or Apathy Responses about Reviews

	Mean value	Standard deviation	Max. Mean value
Sympathy	0.91	1.99	596.8
Apathy	0.42	0.80	475.8

Table 2. Terminologies

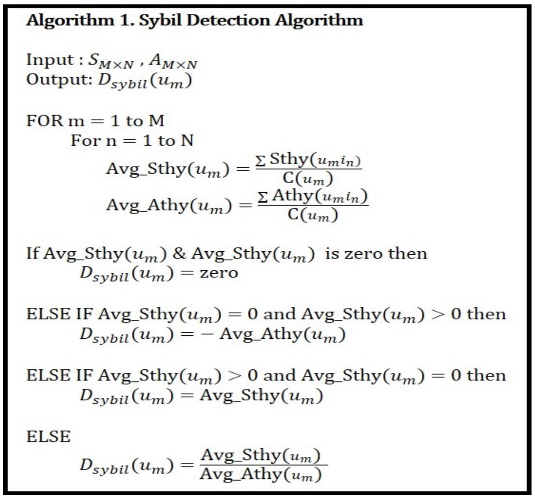
Notation	Description
M	The number of users
N	The number of items
u_m	m^{th} user
i_n	n^{th} item
$C(u_m)$	Count of u_m reviews
R_{MxN}	Rating matrix consist of user m and item n
S_{MxN}	Sympathy matrix consist of user m and item n
A_{MxN}	Apathy matrix consist of user m and item n
$Sthy(u_m i_n)$	u_m's sympathy count on item n
$Athy(u_m i_n)$	u_m's apathy count on item n
Avg_Sthy(u_m)	Average of sympathy count on user m
Avg_Athy(u_m)	Average of apathy count on user m
$D_{sybil}(u_m)$	Decision value of user m is being Sybil

1. Plus, Table 2 shows terminologies in this paper.

According to the results of data analysis, users wrote 13.33 reviews in average, and each review got an average of 1.10 sympathy and 0.49 apathy responses. Through this, we could confirm that the responses were extremely divided. Based on the results, we tried to enhance the performance of item recommendation for normal users by using the mean values of the concerned users.

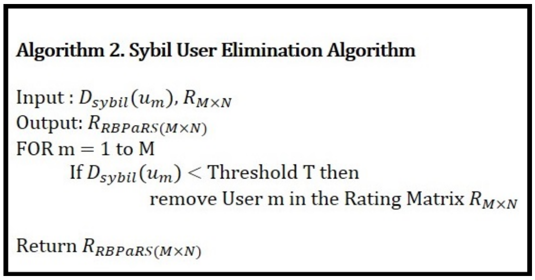
3.2 Algorithm for Detection of Sybil (Abnormal) Users by Applying Sympathy and Apathy Responses

The algorithm to decide Sybil users is as shown in Fig. 1. First, the calculation of each user's average sympathy and apathy ratings and figuring out the ratio of average sympathy and apathy. Plus, in Fig. 2 shows algorithm for removing Sybil user from rating matrix. It is called Robust BiPArtite Recommender System(RBPaRS).



```
Algorithm 1. Sybil Detection Algorithm

Input : $S_{M\times N}$ , $A_{M\times N}$
Output: $D_{sybil}(u_m)$

FOR m = 1 to M
        For n = 1 to N
                Avg_Sthy($u_m$) = $\frac{\Sigma \text{Sthy}(u_m i_n)}{\text{C}(u_m)}$
                Avg_Athy($u_m$) = $\frac{\Sigma \text{Athy}(u_m i_n)}{\text{C}(u_m)}$

If Avg_Sthy($u_m$) & Avg_Sthy($u_m$) is zero then
        $D_{sybil}(u_m)$ = zero

ELSE IF Avg_Sthy($u_m$) = 0 and Avg_Sthy($u_m$) > 0 then
        $D_{sybil}(u_m)$ = − Avg_Athy($u_m$)

ELSE IF Avg_Sthy($u_m$) > 0 and Avg_Sthy($u_m$) = 0 then
        $D_{sybil}(u_m)$ = Avg_Sthy($u_m$)

ELSE
        $D_{sybil}(u_m) = \frac{\text{Avg\_Sthy}(u_m)}{\text{Avg\_Athy}(u_m)}$
```

Fig. 1. Sybil Detection Algorithm



```
Algorithm 2. Sybil User Elimination Algorithm

Input : $D_{sybil}(u_m)$, $R_{M\times N}$
Output: $R_{RBPaRS(M\times N)}$
FOR m = 1 to M
        If $D_{sybil}(u_m)$ < Threshold T then
                remove User m in the Rating Matrix $R_{M\times N}$

Return $R_{RBPaRS(M\times N)}$
```

Fig. 2. Sybil User Elimination Algorithm

The average sympathy and apathy values of each user were calculated, and the concerned sympathy average was divided with the apathy average, which was used to divide into normal and Sybil users. When this value was greater than the threshold, it was considered a normal user, or when it was smaller than the threshold, it was regarded as a Sybil user. When predicting item in Chapter 3.3, we formed the matrices by excluding the reviews written by the concerned user.

3.3 Matrix Factorization

We used the model-based MF technique for recommendation of items by our proposed recommender system for the remaining users after excluding Sybil users. After forming a matrix by mapping users and items to the latent feature and the dimension respectively, the MF multiplied the former by the latter to create the user-item matrices.

$$R \approx U^T V \quad (1)$$

$U \in \mathbb{R}^{k \times m}$ and $V \in \mathbb{R}^{k \times n}$ are the matrices for users and items respectively; k is the dimension of the R matrix [k<min(m, n)]. The above equation can be converted into a question of the least-squares minimization of the Singular Value Decomposition (SVD), which was expressed as $\min \| R - U^T V \|_F$. The indicator function, which computed only those which contained the actual ratings given by users in comparison with the dimension of the R matrix, was expressed as I_{ij} when the user (i) gives a rating to the item, I_{ij} is 1, or if not, $I_{ij}=0$. This can be expressed as follows.

$$\min_{U,V} L(R,U,V) = \frac{1}{2}\sum_i^m \sum_j^n I_{ij}(R_{ij} - U_i^T V_j)^2 \quad (2)$$

In this paper, we used the gradient descent method to solve the optimization problem. Also, in order to prevent the overfitting problem, Equation (2) can be expressed as Equation (3) by including the normalization part.

$$\min_{U,V} L(R,U,V) = \frac{1}{2}\sum_i^m \sum_j^n I_{ij}(R_{ij} - U_i^T V_j)^2 + \frac{\lambda_1}{2}\|U\|_F^2 + \frac{\lambda_2}{2}\|V\|_F^2 \quad (3)$$

Herein, F means the Frobenius norm.

IV. Performance evaluation

4.1 Dataset

For this experiment, we collected users′ ratings and reviews about those movies which were released for 5 years from 2009 and 2013 from Naver movie webpage (movie.naver.com). We selected those users who wrote more than 5 movie reviews and who posted more than 10 reviews, which were summarized as seen in table 3.

Table 3. Data Set

Name	#user	#items	#ratings	Scale
Naver Movie	49,543	2,177	659,794	{1,2,,,10}

We evaluate the proposed scheme using both synthetic and real-world movie site of Korea. We show that the proposed scheme is able to accurately identify Sybil with low false positive rates and low false negative rates. The proposed scheme is resilient to noise in our prior knowledge about known legitimate IDs and Sybils. Moreover, the proposed scheme performs orders of magnitudes better than existing Sybil classification mechanisms and significantly better than existing Sybil ranking mechanisms.

4.2 Performance Evaluation

To compare the performances between our proposed scheme and the existing method, we used the Mean Absolute Error (MAE), which is widely used as the standard indicator of accuracy for recommendation systems.

$$\mathrm{MAE} = \frac{1}{T}\sum_{i,j} |R_{i,j} - R'_{i,j}| \quad (4)$$

R_{ij} is the rating about the item (j) given by the user (i); R'_{ij} is the predicted value estimated through the MF; T is the number of data included in the test data. It can be said that a lower value means that the actual value is closer to the predicted value.

4.3 Experimental results

We classified 90% of the collected data as the training group and the remaining 10% as the test group, and conducted each experiment for five times to produce the mean values. We experiment various sybil classification thresholds T between －0.3 to 0.

In Fig. 3, when we regarded values lower than the threshold value of －0.05 as ‘Sybil’, we could obtain the highest level of accuracy. Although the existing MF method could not respond to Sybil attacks, our proposed recommender system showed a relatively excellent performance, because it removed all ratings which were regarded as Sybil attacks. Also, through the results, we could confirm the possible influence of Sybil users on the reliability of sympathy and apathy responses.

To verify our scheme's utility, we combine our scheme with STA [8] and measure the MAE value in Fig. 4. The union of STA with our scheme have best accuracy results. These two schemes are complementary cooperation of finding sybil users. It means using user's bipartite information can efficiently adapt others and enhance the accuracy performance.

	Basic MF	RBPaRS, T: -0.3	RBPaRS, T: -0.2
MAE	1.878643	1.8715	1.865543

	RBPaRS, T: -0.1	RBPaRS, T: -0.05	RBPaRS, T: 0
MAE	1.858678	**1.849535**	1.865133

Fig. 3. MAE comparison between basic MF and the proposed scheme

	RBPaRS	STA	STA+RBPaRS
MAE	1.849535	1.795198	1.769765

Fig. 4. Result of combination of our scheme and STA

4.4 Sybil Threshold Analysis

We study the threshold of sybil classification. In Naver movie dataset, and why sybil threshold －0.05 is the best performance in our experiments.

First of all, we classify the users by using $D_{sybil}(u_m)$ values. Moreover, to verify our algorithm, we exploit the STA[8] algorithm and using the STA classification values. If user's STA value is close to 1, then user is classified as sybil. In Fig. 5, we show that users who have low $D_{sybil}(u_m)$ values not only get no sympathy values but also higher STA values than users who have high $D_{sybil}(u_m)$ values. Users who get many sympathy value have relatively low STA values. So, it is meaningful sybil threshold in our experiment.

	Number of users	Sympathy Average	Sympathy Average	STA value Average
T≤ −0.05	3,313	0	0.238	0.2278
T> −0.05	46,230	3.2524	0.9109	0.2001

Fig. 5. Analysis of sybil threshold value

V. Conclusion

This study could establish a robust recommender system through analysis of sympathy and apathy responses. By the utilizing a review evaluation method, the proposed recommender system is highly likely to be used as a whole new approach compared to the existing rating-based system. The results of this approach outperforms the previous ones and shows the enhanced accuracy performance. Moreover, it would be exploited other robust recommender system.

References

〔1〕 J.R Douceur, “The Sybil attack,” In International Workshop on Peer-to- Peer Systems, pp. 251 - 260, Mar. 2002.

〔2〕 M.P. O’Mahony, N.J. Hurley and G.C.M. Silverstre, “Promoting recommendations: An attack on collaborative filtering,” In Proceedings of the International Conference on Database and Expert Systems Applications. pp. 494 - 503, Sep. 2002.

〔3〕 M.P. O’Mahony, N.J. Hurley, G.C.M. Silverstre, “Recommender systems: Attack types and strategies,” In Proceedings of the 20st National Conference on Artificial Intelligence, pp. 334 - 339, Jul. 2005.

〔4〕 B. Mobasher, R. Burke, and JJ. Sandvig, “Model-based collaborative filtering as a defense against profile injection attacks,” In Proceedings of the 21st National Conference on Artificial Intelligence, pp. 1388-1393, Jul. 2006.

〔5〕 R. Burke, B. Mobasher, C.A. Williams and R. Bhaumik. “Classification features for attack detection in collaborative recommender systems,” In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 542-547, Aug. 2006.

〔6〕 N. Hurley, Z. Cheng, and M. Zhang. “Statistical attack detection,” In Proceedings of the third ACM conference on Recommender systems. pp. 149-156, Oct. 2009.

〔7〕 C.A. Williams, B. Mobasher and R. Burke. “Defending Recommender Systems: Detection of Profile Injection Attacks,” Service Oriented Computing and Applications, vol. 1, no. 3, pp. 157-170. Nov. 2007.

〔8〕 T. Noh H. Oh, G. Noh and C. Kim. “STA : Sybil Type-aware Robust Recommender System,” KIISE Transactions on Computing Practices, vol. 21, no. 10, pp. 670-679, Oct. 2015.

〔9〕 Z. Zhang and S.R Kulkarni, “Detection of shilling attacks in recommender systems via spectral clustering”, In Proceedings of the International Conference on Information Fusion, pp. 1-8, Jul. 2014.

〔10〕 N.Z. Gong, M. Frank, and P. Mittal, “SybilBelief: A Semi-Supervised Learning Approach for Structure-Based Sybil Detection,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 6, pp. 976-987, Jun. 2014.

〈저자소개〉

이 재 훈 (Jaehoon Lee) 학생회원
2011년 2월: 동국대학교 정보통신공학과 졸업
2013년 2월: 서울대학교 컴퓨터공학부 석사
2013년 3월~현재: 서울대학교 컴퓨터공학부 박사과정
〈관심분야〉 정보 보호, 추천 시스템, Sybil attack

오 하 영 (Hayoung Oh) 정회원
2002년 2월: 덕성여자대학교 컴퓨터공학부 졸업
2006년 2월: 이화여자대학교 컴퓨터공학 석사
2013년 2월: 서울대학교 컴퓨터공학부 박사
2016년 8월: 숭실대학교 정보통신전자공학부 조교수
2016년 9월~현재: 아주대학교 다산학부대학 조교수
〈관심분야〉 소셜 정보망, 추천시스템, 무선 네트워크 및 비디오 스트리밍

김 종 권 (Chong-kwon Kim) 정회원
1981년 2월: 서울대학교 산업공학과 졸업
1982년 8월: 미국 조지아공대 Operations Research 석사
1987년 8월: 미국 일리노이 대학교 전산학과 박사
1991년~현재 서울대학교 컴퓨터공학부 교수
〈관심분야〉 무선통신, 이동통신, 추천 시스템, 소셜 네트워크 분석, 성능평가, 네트워크 보안