

Defending Suspected Users by Utilizing Specific Distance Metric in Collaborative Filtering Recommender Systems

Dipeeka Wattamwar, Pallavi Shatalwar, Prachi Yadav, Gauri Nerkar, Prof. B.S.Satpute
Department of Computer Engineering, Dr.D.Y.Patil Institute of Technology, Savitribai Phule Pune University,
Pune, India.

ABSTRACT: Recommender system is an imperative part of the data and internet business biological system. Collaborative filtering (CF) is a vital and well known innovation for recommender system. Collaborative filtering recommender systems (CFRSs) are basic parts of existing well known online business sites to make personalized recommendations. Be that as it may, current CF strategies experience the ill effects of such issues as "shilling" attacks or "profile injection" attacks because of its openness. While an extensive variety of recognition systems have been utilized, some of them depended on calculating similarity between users (consists of attackers and genuine users) keeping in mind the end goal to segregate those attackers. In practice, it is hard to catch every concerned attacker by abusing the similarity of users, in spite of the fact that it can be useful to filtering out more genuine users. Calculating directly the similarity between users on a whole dataset consumed high computation time although they can be powerful to catch the concerned attackers in some degree. In this paper we propose an unsupervised method to detect such attacks. At the first stage, we filter out a part of genuine users in order to reduce the enumeration time. At the second stage, we mainly focus on the effective similarity metric to better differ between attackers and genuine users based on the remaining users of the first stage and modulates the traditional similarity metric and the linkage information between users to improve the accuracy of similarity of users.

KEYWORDS: Recommender system, Shilling attack, Attack detection, Collaborative Filtering (CF), Similarity metric.

I. INTRODUCTION

Background:

Personalization recommender systems (RSs) become more and more popular in e-commerce websites to automatically make personalized suggestions of services or products to customers. Collaborative filtering (CF) is an important and popular technology for recommender systems. There has been a lot of work done both in industry and academia. These methods are classified into user-based CF and item-based CF. The basic idea of user-based CF approach is to find out a set of users who have similar favor patterns to a given user (i.e., "neighbors" of the user) and recommend to the user those items that other users in the same set like, while the item-based CF approach aims to provide a user with the recommendation on an item based on the other items with high correlations (i.e., "neighbors" of the item). In all collaborative filtering methods, it is a significant step to find users' (or items') neighbors, that is, a set of similar users (or items). However, Collaborative filtering recommender systems (CFRSs) are prone to manipulation from attackers due to its openness, which carefully inject chosen attack profiles into CFRSs to bias the recommendation results to their benefits or decrease the trustworthiness of recommendation. These phenomenons are often called "shilling" attacks or "profile injection" attacks. Therefore, constructing an effective detection method to detect the attackers and remove them from the CFRSs is crucial. A number of detection methods have been proposed to make CFRSs resistant to such attacks. Some of them were based on calculating similarity between users (consists of attackers and genuine users) in order to discriminate those attackers. In practice it is difficult to capture all concerned attackers by exploiting the

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

similarity of users, although it can be helpful to filtering out more genuine users and show high computation time, although they can be effective to capture the concerned attackers in some extent. To resolve all the issues in existing systems we propose a new detection method called defending suspected users by exploiting specific distance metric in collaborative filtering recommender systems to make CFRSs resistant to such attacks, which exploits a novel metric for calculating similarity between users. To reduce the computation time, we firstly filter out a part of genuine users in order to reduce the computation time. Since the attackers will target one or more specific items with lowest or highest rating many times if they want to demote (called nuke attack) or promote (called push attack) the items to the recommendation list, we can find out all suspected target items by using an absolute count threshold. At the second stage, we mainly focus on the effective similarity metric to better distinguish between attackers and genuine users based on the remaining users of the first stage.

Motivation:

Collaborative filtering recommender systems (CFRSs) are highly vulnerable to “shilling” attacks or “profile injection” attacks due to its openness. While a wide range of detection techniques have been used, some of them were based on calculating similarity between users (consists of attackers and genuine users) in order to discriminate those attackers. In practice, it is difficult to capture all concerned attackers by exploiting the similarity of users, although it can be helpful to filtering out more genuine users. Moreover, calculating directly the similarity between users on a whole dataset consumed high computation time. Therefore to resolve all above problem we propose a new detection method to make CFRSs resistant to such attacks, which exploits a novel metric for calculating similarity between users.

Objective and Goal:

- To improve accuracy in discriminating attack user form genuine user by proposing the detection algorithm.
- To develop a scheme for outsourced data that takes into account both the security and performance.

II. LITERATURE SURVEY

Number	Paper Name	Author Name	Proposed System	Referred Point
1.	A novel approach to filter out malicious rating profiles from recommender systems	C. Chung, P. Hsu, and S. Huang	This paper proposes novel detection method to make recommender systems resistant to the “shilling” attacks or “profile injection” attacks.	From this Paper, we have referred problem as finding a mapping model between rating behavior and item distribution by exploiting the least-squares approximate solution.
2.	HHT-SVM: An online method for detecting profile injection attacks in collaborative recommender systems	F. Zhang and Q. Zhou	This paper propose online method (called HHT-SVM) to detect profile injection attacks by combining Hilbert-Huang transform (HHT) and support vector machine (SVM), which can work incrementally.	From this Paper, we have referred the empirical mode decomposition (EMD) approach.
3.	Detection of abnormal profiles on group attacks in recommender systems	W. Zhou, Y. S. Koh, J. H. Wen, S. Burki, and G. Dobbie	This paper proposes the FIM technique to generate candidate groups, and then present several features for	From this Paper, we have referred the FIM technique.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

			modeling the abnormal behavior at the group level.	
4.	Defending recommender systems by influence analysis	M. Morid and M. Shajari	This paper proposes an attack detection method based on user influence in recommender systems.	From this Paper, we have referred an attack detection method.
5.	Detection of shilling attacks in recommender systems via spectral clustering	Z. Zhang and S. R. Kulkarni	This paper defines the issue as finding a most extreme submatrix in the user-user similarity matrix.	From this Paper, We have referred a spectral clustering algorithm to find the min-cut solution to estimate the highly correlated group.
6.	Shilling attacks against recommender systems: A comprehensive survey	I. Gunes, C. Kaleli, A. Bilge, and H. Polat	This paper proposes decoding shilling attack detection schemes in detail and robust algorithms proposed so far might open a lead to develop new detection schemes and increase such robust algorithms	From this Paper, We have referred robust recommendation algorithms.
7.	Graph-based detection of shilling attacks in recommender systems	Z. Zhang and S. Kulkarni	This paper presents a method to make recommender systems obstructive to these attacks in the case that the attack profiles are highly correlated with each other.	From this Paper, We have referred how find a maximum submatrix in the similarity matrix.
8.	A survey on shilling attack models and detection techniques for recommender systems	Z. A. Wu, Y. Q. Wang, and J. Cao	This paper reviews the states of art and the main problems of existing works related to shilling attack models and detection techniques, and attempts to sketch an extensive and evident outline for this new and active research realm.	From this Paper, We have referred the shilling attack models and detection techniques.
9.	Attacking item-based recommender systems with power items	C. E. Seminario and D. C. Wilson	This paper shows that the Power Item Attack (PIA) is ready to affect not only user-based and SVD-based recommenders but also the heretofore highly robust item-based approach, using a novel multi-target attack	From this Paper, We have referred the robust item-based approach.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

			vector.	
10.	Pair wised specific distance learning from physical linkages	J. Hu, D. C. Zhan, X. Wu, Y. Jiang, and Z. H. Zhou	This approach exploits the structures of physical linkages and in particular captures the key observations that nonmetric and clique linkages imply the appearance of different or unique semantics, respectively.	From this Paper, We have referred PSD for multi-class learning and further extend it to multi-label learning.

III. SOFTWARE REQUIREMENT SPECIFICATION

User Classes and Characteristics

To design products that satisfy their target users, a deeper understanding is needed of their user characteristics and product properties in development related to unexpected problems that the user's faces every now and then while developing a project. The study will lead to an interaction model that provides an overview of the interaction between user characters and the classes. It discovers both positive and negative patterns in text documents as higher level features and deploys them over low-level features (terms). In proposed work is designed to implement above software requirement. To implement this design following software requirements and hardware requirements are used.

Software Requirements

- Operating System - Windows XP/7
- Programming Language - Java/J2EE
- Software Version - JDK 1.7 or above
- Tools - Eclipse
- Front End - JSP
- Database - Mysql

Hardware Requirements

- Processor - Pentium IV/Intel I3 core
- Speed - 1.1 GHz
- RAM - 512 MB (min)
- Hard Disk - 20GB
- Keyboard - Standard Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - LED Monitor

IV. IMPLEMENTATION STATUS

Our approach consists of two stages: the stage of filtering out genuine users by exploiting suspected target items and the stage of discriminating attackers by employing a new similarity metric. At the first stage, we filter out a part of genuine users in order to reduce the computation time. At the second stage, we mainly focus on the effective similarity metric to better distinguish between attackers and genuine users based on the remaining users of the first stage. To reduce the computation time for our proposed method, we firstly make a preprocessing to filter out more genuine users as far as possible. As is known, the attackers should promote or demote one or more target items with the highest or

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

lowest rating to achieve their attack intentions. In other words, the attackers will target these specific items many times if they want to push or nuke the items to the recommendation list. To determine the target items, we design an absolute count threshold ε which pushes or nukes the same items with the highest or lowest at least ε times. If the count for an item is greater than ε , then the item *susi* is regarded as suspected target item. Users (consist of attackers and genuine users) who rated the *susi* with the highest *rmax* or lowest *rmin* are considered as attackers. It is noteworthy that there will be more false positives or false negative if ε is too small or large. Based on the remaining result of first stage, we continue to filter out all genuine users as far as possible by employing a new similarity metric. Since traditional similarity metrics such as PCC, Cosine similarity etc. are difficult to fully measure the similarity between users, we exploit a novel similarity metric to distinguish between attackers and genuine users, which combines the global and local similarity metrics and uses the linkage information between users to further improve accuracy of similarity measurement. Inspired from the idea of pair wised specific distance, we also constrain the distance between users from the same class to be smaller than their Euclidean distance, and the distance between users from different classes larger than their Euclidean distance.

The System consist of following modules,

- Authentication
- Detect Suspected Item
- Detect Suspected Users
- Recommendation

V. COMPARISON BETWEEN EXISTING SYSTEM AND PROPOSED SYSTEM

In existing system, constructing an effective detection method to detect the attackers and remove them from the CFRSs is crucial. In the existing systems a number of detection methods have been proposed to make CFRSs resistant to such “shilling” attacks or “profile injection” attacks. Some of them were based on calculating similarity between users (consists of attackers and genuine users) in order to distinguish those attackers. Practically speaking it is hard to catch all concerned attackers by exploiting the similarity of users, although it can be helpful to filtering out more genuine users and attackers and show high computation time, although these existing systems can be effective to capture the concerned attackers in some extent.

The proposed system proposes a new detection method to make CFRSs resistant to such attacks, which exploits a novel metric for calculating similarity between users. To lower the computation time, firstly we filter out more genuine users as far as possible determined by using mistrusted target items. Since the attackers will target one or more specific items with lowest or highest rating many times if they want to demote (called nuke attack) or promote (called push attack) the items to the recommendation list, we can find out all mistrusted target items by using an complete count threshold. Based on the remaining users, we employ a new similarity metric inspired from the pair wised specific distance, directing to measure impressively the similarity between users.

VI. ALGORITHM FOR RELEVANT FEATURE DISCOVERY

- **Detection Algorithm:**

Input: Data matrix **M**;
Empirical thresholds ε and τ ;
Termination threshold δ ;
Maximum iteration limit T ;
Linkage graph for the whole dataset G ;

Output: The detected result **DR**;

Process:

Step 1:

```

1:  $\{(i, N_i)\}_i^{|I|} = \{(i, N_i) | N_i = \text{the number of ratings on}$   

 $\text{item } i \text{ with } r_{min}\}$ ;  

2:  $item_{sus} = null$ ;  

3:  $user_{sus} = null$ ;  

4: if  $N_i > \varepsilon$  then  

5:    $item_{sus} \leftarrow i$ ;  

6: end if  

7: for each user  $u$  in  $U$  do  

8:   if  $u$  rated item  $i$  ( $i \in item_{sus}$ ) with  $r_{min}$  then  

9:     add  $u$  to  $user_{sus}$ ;  

10:  end if  

11: end for

```

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Step 2:

```

12: Initialize  $w^{(0)} = 1$ ;
13: for  $t = 1, \dots, T$  do
14:   for  $wv \in G \wedge u, v \in user_{sus}$  do
15:     Learn  $w_{uv}$  by solving Eq. 3;
16:   end for
17:   if  $\|w^{(t)} - w^{(t-1)}\|_2 \in \delta$  then
18:     break;
19:   end if
20: end for
21:  $w^{(0)} = w$ ;
22: Calculate similarity between remained users in
 $user_{sus}$  by using the new metric;
23: for each pair of users  $u$  and  $v$  in  $user_{sus}$  do
24:   if  $s_{u,v} > \tau$  then
25:     remain  $u$  and  $v$  in  $user_{sus}$ ;
26:   else
27:     remove  $u$  and  $v$  from  $user_{sus}$ ;
28:   end if
29: end for
30:  $DR = user_{sus}$ ;
31: Return: The detected result  $DR$ ;

```

- **Collaborative Filtering Technique:**

Collaborative filtering, also referred to as social filtering, filters information by using the recommendations of other people. It is based on the idea that people who agreed in their evaluation of certain items in the past are likely to agree again in the future. A person who wants to see a movie for example, might ask for recommendations from friends. The recommendations of some friends who have similar interests are trusted more than recommendations from others. This information is used in the decision on which movie to see.

We uses,

- **Item-based collaborative filtering**
- **User-based collaborative filtering**

The basic idea of user-based CF approach is to find out a set of users who have similar favor patterns to a given user (i.e., “neighbors” of the user) and recommend to the user those items that other users in the same set like, while the item-based CF approach aims to provide a user with the recommendation on an item based on the other items with high correlations (i.e., “neighbors” of the item).

VII. SYSTEM ARCHITECTURE

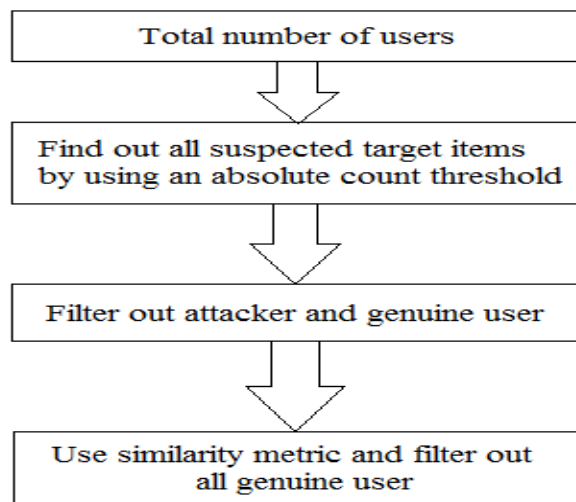


Figure 1: System Architecture

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

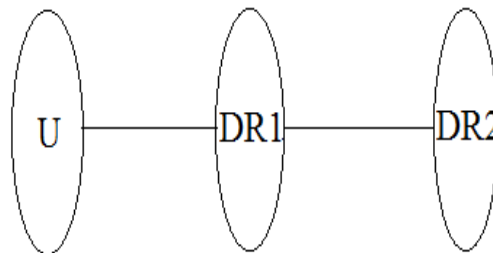
Vol. 7, Issue 5, May 2018

Our proposed approach consists of two stages: the stage of filtering out genuine users by misusing suspected target items and the stage of separating attackers by employing a new similarity metric. At the first stage, we filter out a part of genuine users in order to decrease the computation time. To decide the target items, we design an absolute count threshold ε which pushes or nukes the same items with the highest or lowest at least ε times. If the count for an item is greater than ε , then the item *susi* is regarded as suspected target item. Users (consist of attackers and genuine users) who rated the *susi* with the highest *rmax* or lowest *rmin* are considered as attackers. It is noticeable that there will be more false positives or false negative if ε is too small or large. At the second stage, we mainly focus on the effective similarity metric to better differentiate between attackers and genuine users based on the remaining users of the first stage.

IX. MATHEMATICAL MODULE

The following terms shows in detail working of project.

A) Mapping Diagram



Where,

U = Total users

DR1 = First stage detecting result

DR2 = Second stage detecting result

B) Set Theory

$S = \{s, e, X, Y, \}$

Where,

s = Start of the program.

1. Register with system.

2. Login with web page.

X = Input of the program.

Collaborative filtering recommender systems

Y = Output of the program.

Defended Suspected Users

e = End of the program.

Detection rate is defined as the number of detected attack profiles divided by the number of attack profiles.

Detection rate = No of Detection / No of Attack Profiles

False alarm rate is the number of genuine profiles that are predicted as attack profiles divided by the number of genuine profiles.

False alarm rate = No of False alarm / No of Genuine Profiles

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

- **Space Complexity:**

The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

- **Time Complexity:**

Check No. of patterns available in the datasets= n

If (n>1) then retrieving of information can be time consuming. So the time complexity of this algorithm is O.
= Failures and Success conditions.

- **Failures:**

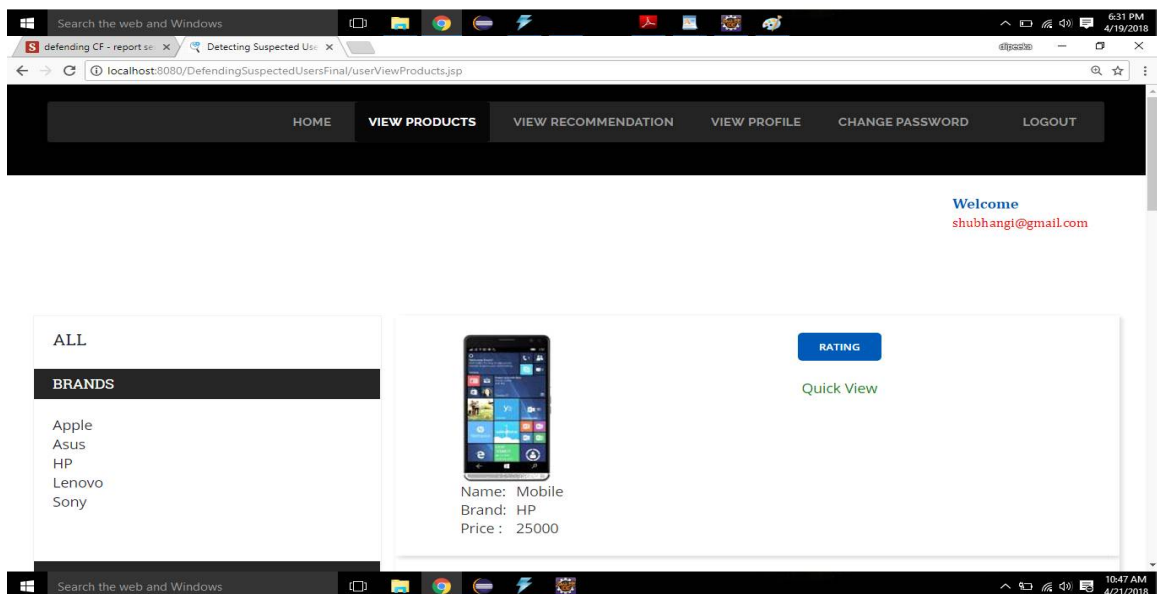
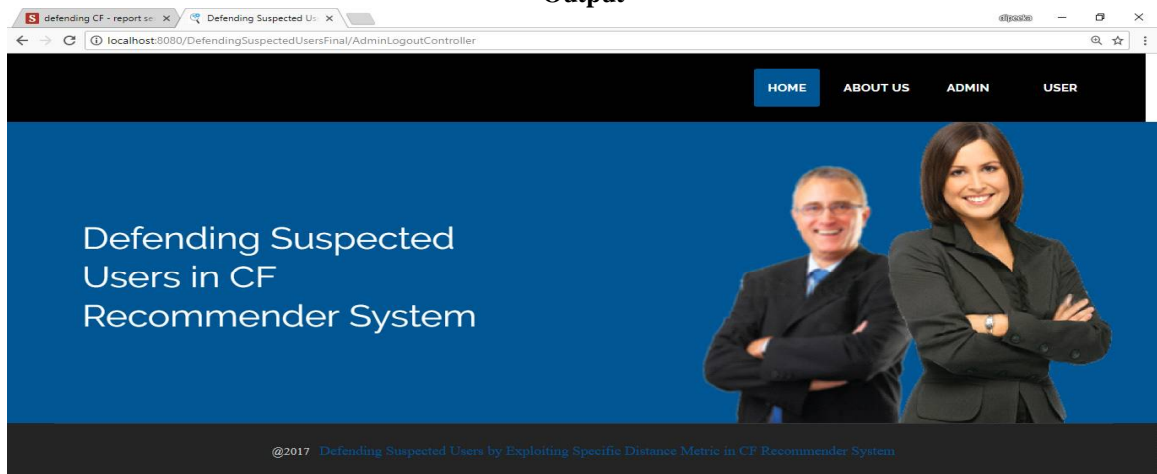
Due to schilling attack recommendation lists can be altered by fake users.

- **Success:**

Fake profile injection can be stopped and we can provide unbiased recommendation to user.

X. EXPERIMENTAL SET UP AND RESULT TABLE

- **Output**

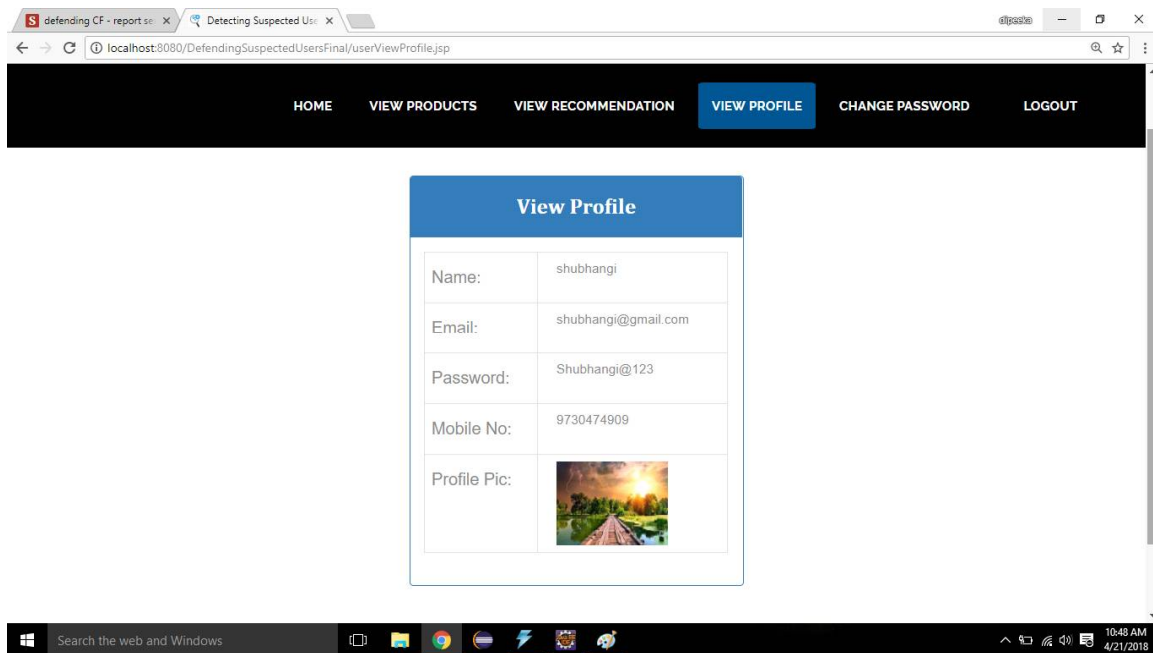


International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018




defending CF - report se X Detecting Suspected Users X

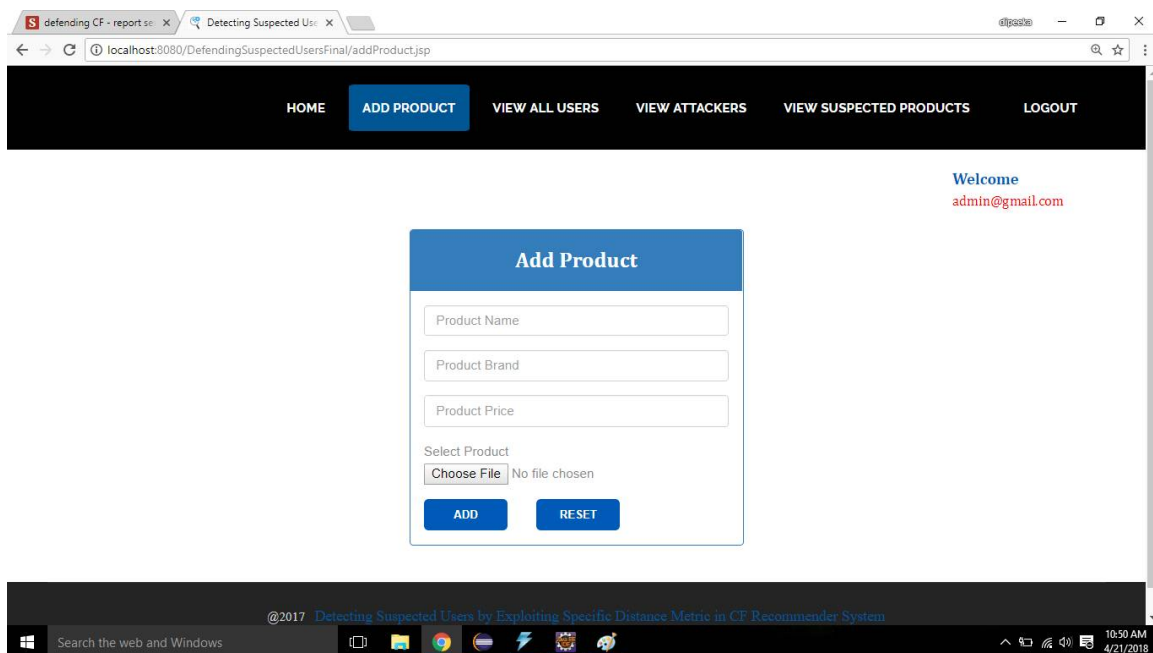
localhost:8080/DefendingSuspectedUsersFinal/userViewProfile.jsp

HOME VIEW PRODUCTS VIEW RECOMMENDATION **VIEW PROFILE** CHANGE PASSWORD LOGOUT

View Profile

Name:	shubhangi
Email:	shubhangi@gmail.com
Password:	Shubhangi@123
Mobile No:	9730474909
Profile Pic:	

Search the web and Windows 10:49 AM 4/21/2018



defending CF - report se X Detecting Suspected Users X

localhost:8080/DefendingSuspectedUsersFinal/addProduct.jsp

HOME **ADD PRODUCT** VIEW ALL USERS VIEW ATTACKERS VIEW SUSPECTED PRODUCTS LOGOUT

Welcome
admin@gmail.com

Add Product

Product Name

Product Brand

Product Price

Select Product
 No file chosen

@2017 Detecting Suspected Users by Exploiting Specific Distance Metric in CF Recommender System

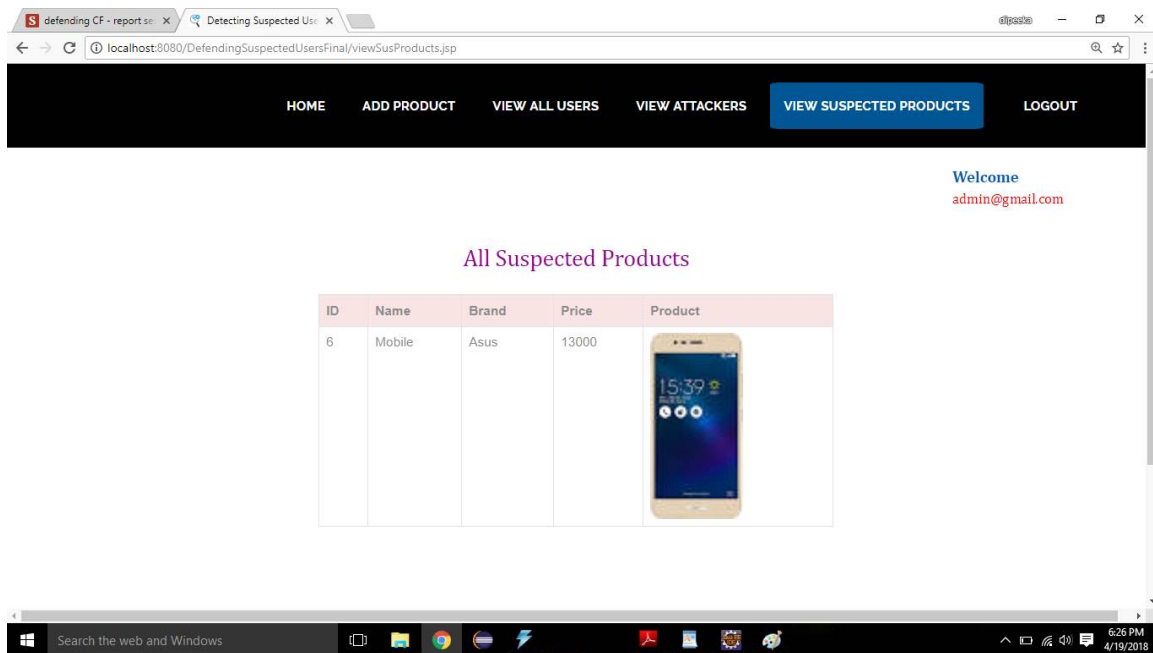
Search the web and Windows 10:50 AM 4/21/2018

International Journal of Innovative Research in Science, Engineering and Technology


(A High Impact Factor, Monthly, Peer Reviewed Journal)

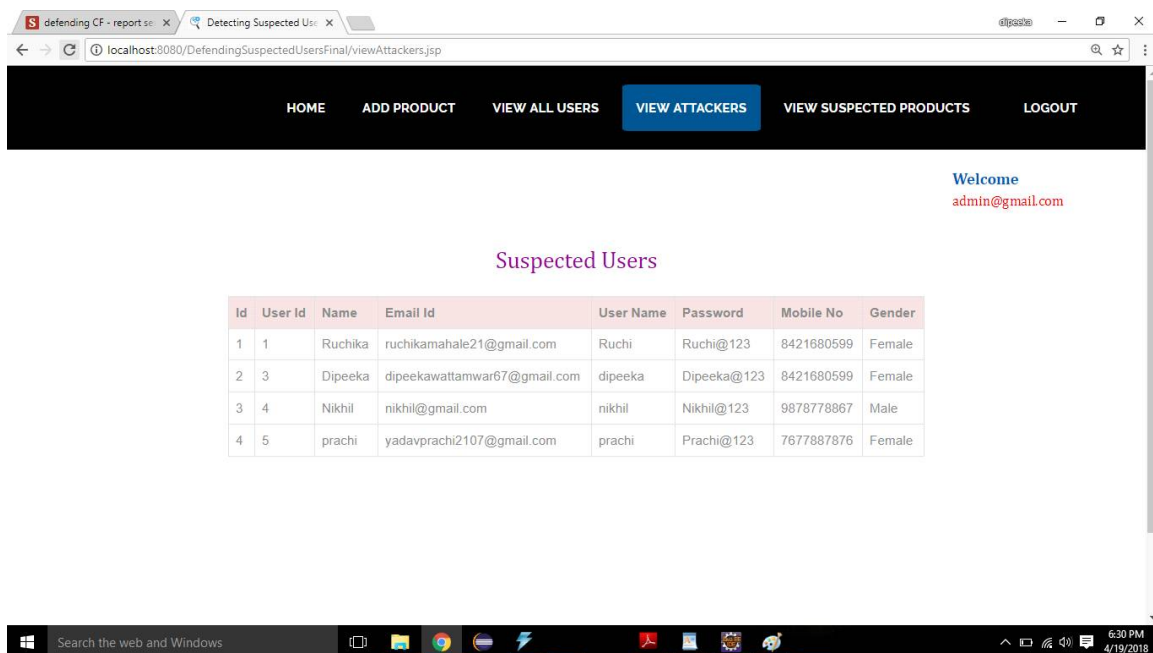
Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018



The screenshot shows a web browser window with the URL `localhost:8080/DefendingSuspectedUsersFinal/viewSusProducts.jsp`. The page has a navigation bar with links: HOME, ADD PRODUCT, VIEW ALL USERS, VIEW ATTACKERS, VIEW SUSPECTED PRODUCTS (highlighted), and LOGOUT. A welcome message for `admin@gmail.com` is displayed. The main content area is titled "All Suspected Products" and contains a table with one row of data:

ID	Name	Brand	Price	Product
6	Mobile	Asus	13000	



The screenshot shows a web browser window with the URL `localhost:8080/DefendingSuspectedUsersFinal/viewAttackers.jsp`. The page has a navigation bar with links: HOME, ADD PRODUCT, VIEW ALL USERS, VIEW ATTACKERS (highlighted), VIEW SUSPECTED PRODUCTS, and LOGOUT. A welcome message for `admin@gmail.com` is displayed. The main content area is titled "Suspected Users" and contains a table with four rows of data:

Id	User Id	Name	Email Id	User Name	Password	Mobile No	Gender
1	1	Ruchika	ruchikamahale21@gmail.com	Ruchi	Ruchi@123	8421680599	Female
2	3	Dipeeka	dipeekawattamwar67@gmail.com	dipeeka	Dipeeka@123	8421680599	Female
3	4	Nikhil	nikhil@gmail.com	nikhil	Nikhil@123	9878778867	Male
4	5	prachi	yadavprachi2107@gmail.com	prachi	Prachi@123	7677887876	Female

XI. CONCLUSION

We have exhibited an unsupervised detection method called defending suspected users by exploiting specific distance metric in collaborative filtering recommender systems for detecting attacks like “shilling” attacks or “profile injection”

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

attacks, which exploits the pair wise specific distance to generate a similarity metric. Firstly, we filter our more genuine users as far as possible determined by using mistrusted target items. And then, based on the remained users, we continue to filter out more genuine users used by an empirical threshold of the new similarity metric.

REFERENCES

1. C. Chung, P. Hsu, and S. Huang, "A novel approach to filter out malicious rating profiles from recommender systems," *Journal of Decision Support Systems*, pp. 314–325, 2013.
2. F. Zhang and Q. Zhou, "HHT-SVM: An online method for detecting profile injection attacks in collaborative recommender systems," *Knowledge-Based Systems*, 2014.
3. W. Zhou, Y. S. Koh, J. H. Wen, S. Burki, and G. Dobbie, "Detection of abnormal profiles on group attacks in recommender systems," *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 955–958, 2014.
4. M. Morid and M. Shajari, "Defending recommender systems by influence analysis," *Information Retrieval*, pp. 137–152, 2014.
5. Z. Zhang and S. R. Kulkarni, "Detection of shilling attacks in recommender systems via spectral clustering," *International Conference on Information Fusion*, pp. 1–8, 2014.
6. Gunes, C. Kaleli, A. Bilge, and H. Polat, "Shilling attacks against recommender systems: A comprehensive survey," *Artificial Intelligence Review*, pp. 1–33, 2012.
7. Z. Zhang and S. Kulkarni, "Graph-based detection of shilling attacks in recommender systems," *IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1–6, 2013.
8. Z. A. Wu, Y. Q. Wang, and J. Cao, "A survey on shilling attack models and detection techniques for recommender systems," *Science China*, vol. 59, no. 7, pp. 551–560, 2014.
9. C. E. Seminario and D. C. Wilson., "Attacking item-based recommender systems with power items," *ACM Conference on Recommender Systems*, pp. 57–64, 2014.
10. J. Hu, D. C. Zhan, X. Wu, Y. Jiang, and Z. H. Zhou, "Pair wise specific distance learning from physical linkages," *ACM Trans. Knowl. Discov. Data*, 2014.