Shill Block

A V S Radhika, Merugu Yamini, Pamu Sri Vaishnavi

[1]Assistant professor Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, India

[2, 3,]Students, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, India.

Abstract

Existing shilling attack detection approaches focus mainly on identifying individual attackers in online recommender systems and rarely address the detection of group shilling attacks in which a group of attackers colludes to bias the output of an online recommender system by injecting fake profiles.

In this article, we propose a group shilling attack detection method based on the bisecting K-means clustering algorithm. First, we extract the rating track of each item and divide the rating tracks to generate candidate groups according to a fixed time interval. Second, we propose item attention degree and user activity to calculate the suspicious degrees of candidate groups.

Finally, we employ the bisecting K-means algorithm to cluster the candidate groups according to their suspicious degrees and obtain the attack groups. The results of experiments on the Netflix and Amazon data sets indicate that the proposed method outperforms the baseline methods.

1-INTRODUCTION

Online recommender systems play a vital role in delivering personalized content, yet they face serious threats from malicious activities known as shilling attacks. While many approaches exist to detect individual attackers, a more sophisticated and damaging form group shilling attacks involves multiple users acting in coordination to manipulate system outcomes. These groups create fake profiles and work together to promote or demote specific items, often bypassing traditional detection methods. This project introduces a novel detection technique using Bisecting K-Means Clustering, which analyzes rating patterns over time, evaluates user engagement through specific behavioral metrics, and groups suspicious profiles for investigation. By focusing on collective behavior rather than isolated actions, the method aims to strengthen the integrity and trustworthiness of recommender systems. This project focuses on detecting group shilling attacks in online recommender systems by identifying coordinated malicious behavior. It involves analyzing user rating patterns within specific time intervals, calculating behavioral metrics like item attention and user activity, and clustering suspicious groups using the Bisecting K-Means algorithm. The approach is designed to work on large-scale datasets, such as those from Netflix and Amazon, and can be applied to any recommendation-based platform to improve accuracy, security, and user trust.

2-REQUIREMENT ANALYSIS

Functional Requirements

Admin Module

The Admin Module provides functionalities for managing the system and monitoring suspicious activities. Admins can log in securely, add new products, view the list of existing products, and delete products as needed. They can view user-submitted reviews and identify potentially malicious patterns. The admin can block users suspected of participating in group shilling attacks. Additionally, they can run the Bisecting K-Means clustering algorithm to detect such attacks and view the identified suspicious user groups. The module also allows the admin to log out securely after use.

User Module

The User Module enables users to interact with the platform. Users can register with their details and log in to access their accounts. They can view available products, buy selected items, and view their transaction history. After making purchases, users can add reviews and ratings to products, which become part of the data analyzed for detecting shilling attacks. Finally, users can securely log out of the system.

Non-Functional Requirements

- Portability : The ability of the detection system to operate seamlessly across various platforms and environments with minimal configuration or adaptation.
- Reliability: The technique must consistently identify group shilling attacks under different scenarios, ensuring minimal false positives and negatives.
- Performance: The detection algorithm should process data efficiently, especially for large-scale systems, without introducing significant computational overhead.
- Security: The method should robustly prevent attackers from manipulating the system, maintaining data integrity, and resisting both known and emerging attack strategies.

3-DESIGN

Architecture

Project architecture represents number of components we are using as a part of our project and the flow of request processing i.e. what components in processing the request and in which order. An architecture description is a formal description and representation of a system organized in a way that supports reasoning about the structure of the system. Architecture is of two types. They are

(1) Software Architecture

(2) Technical Architecture

Software Architecture

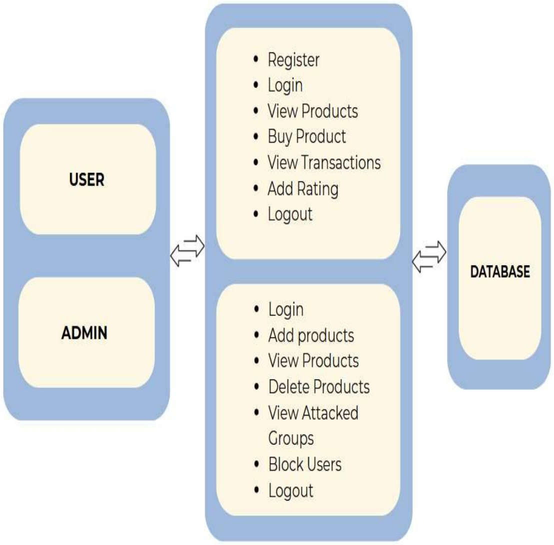


Fig.3.1 Software Architecture

Technical Architecture :

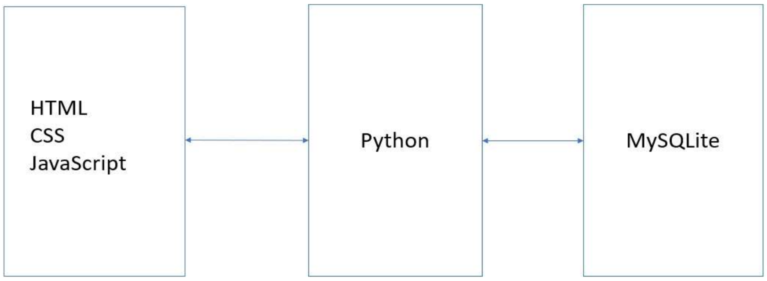


Fig.3.2 Technical Architecture

Data Flow Diagram

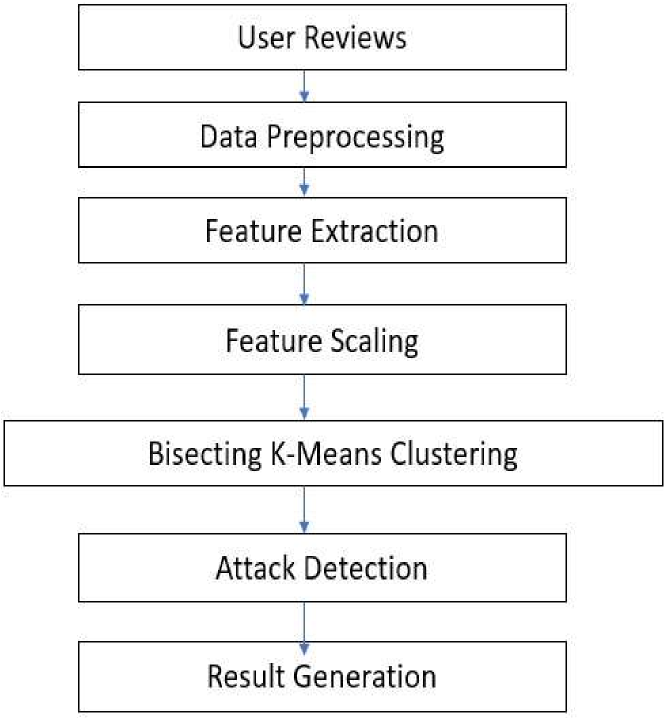


Fig.3.3 Data Flow Diagram

Technologies

4-IMPLEMENTATION

HTML (Hyper Text Markup Language) is the standard language for creating web pages. It structures content using elements like headings, paragraphs, lists, links, and multimedia. Each element is defined by tags, which browsers interpret to render the page. HTML is essential for building websites and is often paired with CSS and JavaScript for styling and interactivity.

CSS

CSS (Cascading Style Sheets) is a powerful language used to style and format HTML elements, enhancing the appearance and user experience of web pages. It controls visual aspects like colors, fonts, spacing, layouts, and responsive designs for different devices. CSS allows separation of content (HTML) from design, making websites more manageable and scalable. Styles can be applied using inline, internal, or external stylesheets, offering flexibility in implementation. CSS supports features like animations, transitions, and media queries for dynamic and interactive designs. Together with HTML and JavaScript, CSS forms the foundation.

JavaScript

JavaScript is a programming language that makes websites interactive and dynamic. While HTML structures a webpage and CSS styles it, JavaScript adds life to it. For example, it enables features like image sliders, interactive forms, pop-ups, animations, and buttons that respond when clicked. It can also update content on the page without reloading, like showing live scores or search suggestions. JavaScript works in web browsers and is also used on servers with tools like Node.js to build full web applications. It's a key part of modern web development, making websites more engaging and user-friendly.

Python

Python is a versatile and widely-used high-level programming language known for its simplicity and readability. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming, making it suitable for a wide range of applications. In the context of data science and machine learning, Python provides powerful libraries such as NumPy, pandas, scikit-learn, and matplotlib, which facilitate efficient data processing, analysis, and visualization. Its ease of use and rich ecosystem make Python an ideal choice for implementing algorithms like Bisecting K-Means for detecting group shilling attacks in online recommendation systems.

MySQLite

MySQLite (commonly referred to as SQLite) is a lightweight, self-contained, and serverless database engine that is widely used for local storage in applications. It is written in C and provides a full-featured SQL database without the need for a separate server process, making it ideal for small to medium-scale projects. SQLite is easy to set up and integrates well with programming languages like Python, allowing developers to perform database operations using simple SQL commands. Its efficiency and minimal configuration requirements make it suitable for managing data in research projects such as detecting group shilling attacks.

5-TESTING

Unit Testing

During this first round of testing, the program is submitted to assessments that focus on specific units or components of the software to determine whether each one is fully functional. The main aim of this endeavor is to determine whether the application functions as designed.

In this phase, a unit can refer to a function, individual program or even a procedure, and White box testing method is usually used to get the job done. One of the biggest benefits of this testing phase is that it can be run every time a piece of code is changed, allowing issues to be resolved as quickly as possible. It quite common for software developers to perform unit tests before delivering software to testers for formal testing.

Integration Testing

Integration testing allows individuals the opportunity to combine all of the units within a program and test them as a group. This testing level is designed to find interface defects between the modules/functions. This is particularly beneficial because it determines how efficiently the units are running together. Keep in mind that no matter how efficiently each unit is running, if they properly integrated, it will affect the functionality of the software program. In order to run these types of tests, individuals can make use of various testing methods, but the specific method that will be used to get the job done will depend greatly on the way in which the units are defined.

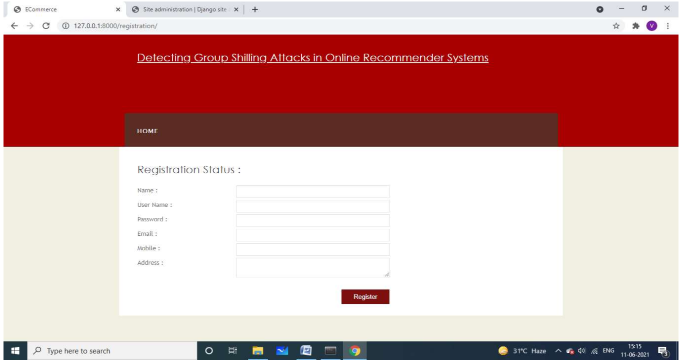
System Testing

System testing is the first level in which the complete application is tested as a whole. The goal at this level is to evaluate whether the system has complied with all of the outlined requirements and to see that it meets Quality Standards. System testing is undertaken by independent testers who haven't played a role in developing the program. This testing is performed in an environment that closely mirrors production. System Testing is very important because it verifies that the application meets the technical, functional, and business requirements that were set by the customer.

Acceptance Testing

The final level, Acceptance testing (or User Acceptance Testing), is conducted to determine whether the system is ready for release. During the Software development life cycle, requirements changes can sometimes be misinterpreted in a fashion that does not meet the intended needs of the users. During this final phase, the user will test the system to find out whether the application meets their business needs. Once this process has been completed and the software has passed, the program will then be delivered to production. The extensiveness of these tests is just another reason why bringing software testers in early is important. When a program is more thoroughly tested, a greater number of bugs will be detected; this ultimately results in higher quality software

6-SCREENSHOTS

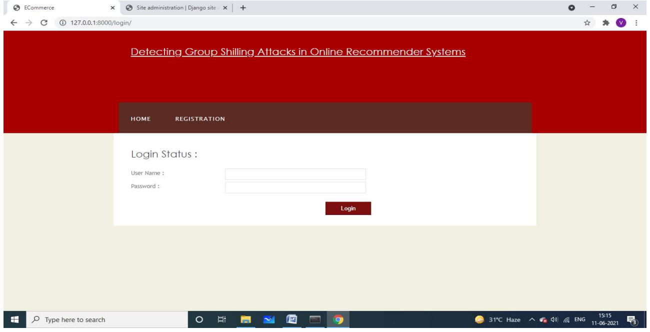


Detecting Group Shilling Attacks in Online Recommender Systems

Registration Status :

Name :
User Name :
Password :
Email :
Mobile :
Address :

Screenshot 6.1 Registration



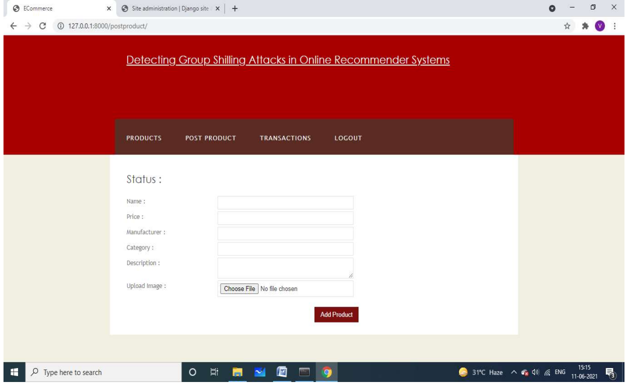
Detecting Group Shilling Attacks in Online Recommender Systems

Login Status :

User Name :

Password :

Screenshot 6.2 Login



Detecting Group Shilling Attacks in Online Recommender Systems

Status :

Name :

Price :

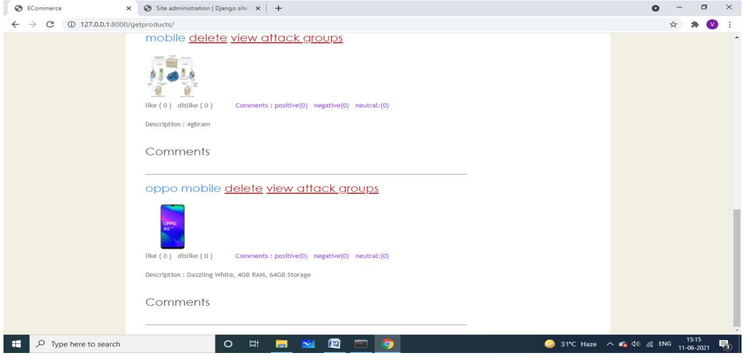
Manufacturer :

Category :

Description :

Upload Image :

Screenshot 6.3 Admin add product


ECommerce
Site administration | Django site
127.0.0.1:8000/getproducts/
Type here to search
31°C Haze
ENG
15:15
11-06-2021

mobile delete view attack groups

like (0) dislike (0) Comments : positive(0) negative(0) neutral:(0)

Description : 4gbram

Comments

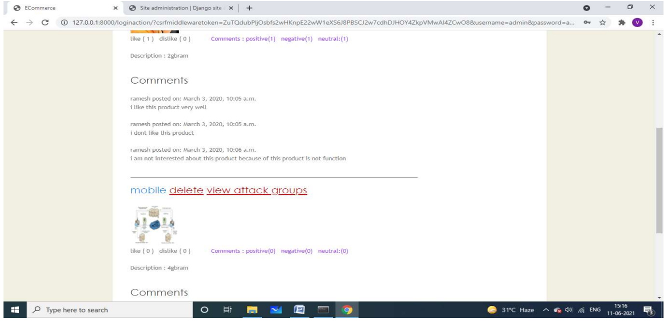
oppo mobile delete view attack groups

like (0) dislike (0) Comments : positive(0) negative(0) neutral:(0)

Description : Dazzling White, 4GB RAM, 64GB Storage

Comments

Screenshot 6.4 Admin View Products


ECommerce
Site administration | Django site
127.0.0.1:8000/loginaction/?csrfmiddlewaretoken=ZuTQdubPljOsbfs2wHKnpE22wW1eXS6J8PBSCJ2w7cdhDJHOY4ZkpVMwAI4ZCwO8&username=admin&password=a...
Type here to search
31°C Haze
ENG
15:16
11-06-2021

like (1) dislike (0) Comments : positive(1) negative(1) neutral:(1)

Description : 2gbram

Comments

ramesh posted on: March 3, 2020, 10:05 a.m.
i like this product very well

ramesh posted on: March 3, 2020, 10:05 a.m.
i dont like this product

ramesh posted on: March 3, 2020, 10:06 a.m.
i am not interested about this product because of this product is not function

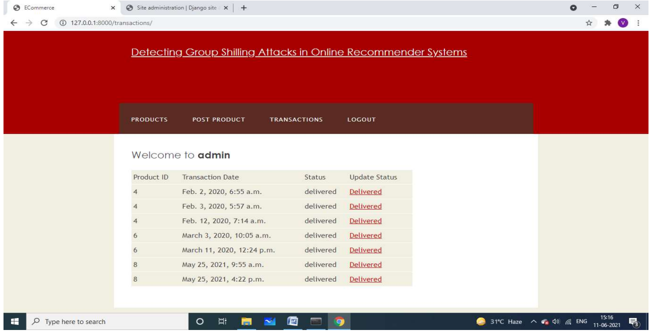
mobile delete view attack groups

like (0) dislike (0) Comments : positive(0) negative(0) neutral:(0)

Description : 4gbram

Comments

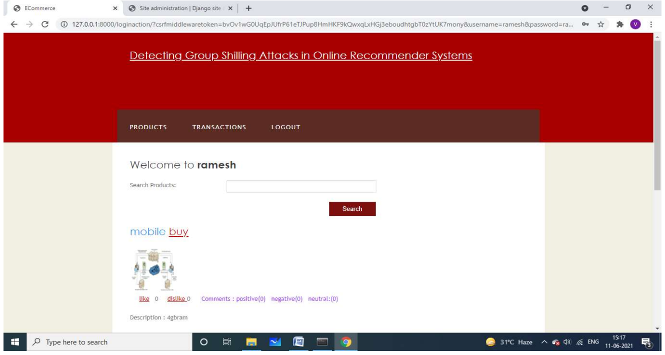
Screenshot 6.5 Admin Manage Products


ECommerce
Site administration | Django site
127.0.0.1:8000/transactions/
Detecting Group Shilling Attacks in Online Recommender Systems
PRODUCTS
POST PRODUCT
TRANSACTIONS
LOGOUT
Type here to search
31°C Haze
ENG
15:16
11-06-2021

Welcome to **admin**

Product ID	Transaction Date	Status	Update Status
4	Feb. 2, 2020, 6:55 a.m.	delivered	Delivered
4	Feb. 3, 2020, 5:57 a.m.	delivered	Delivered
4	Feb. 12, 2020, 7:14 a.m.	delivered	Delivered
6	March 3, 2020, 10:05 a.m.	delivered	Delivered
6	March 11, 2020, 12:24 p.m.	delivered	Delivered
8	May 25, 2021, 9:55 a.m.	delivered	Delivered
8	May 25, 2021, 4:22 p.m.	delivered	Delivered

Screenshot 6.6 Admin View Transactions


ECommerce
Site administration | Django site
127.0.0.1:8000/loginaction/?csrfmiddlewaretoken=bvOv1wG0UqEpJUfrP61eTJPup8HmHKF9kQwxqLxHGj3eboudhtgbT0zYtUK7mony&username=ramesh&password=ra...
Detecting Group Shilling Attacks in Online Recommender Systems
PRODUCTS
TRANSACTIONS
LOGOUT
Search
like
dislike
Type here to search
31°C Haze
ENG
15:17
11-06-2021

Welcome to **ramesh**

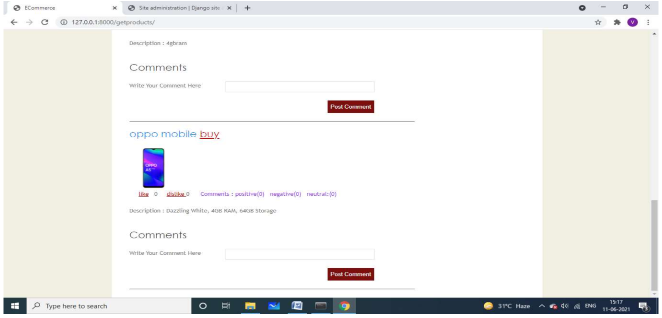
Search Products:

mobile buy

like 0 dislike 0 Comments : positive(0) negative(0) neutral:(0)

Description : 4gbram

Screenshot 6.7 User Search Products



Description : 4gbram

Comments

Write Your Comment Here

oppo mobile buy

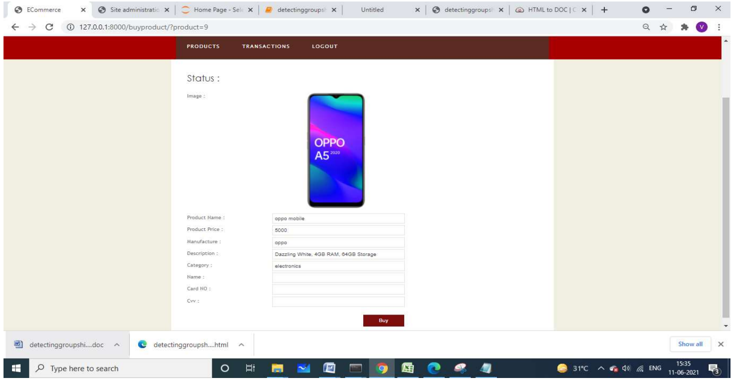
like 0 dislike 0 Comments : positive(0) negative(0) neutral:(0)

Description : Dazzling White, 4GB RAM, 64GB Storage

Comments

Write Your Comment Here

Screenshot 6.8 User Buy Products

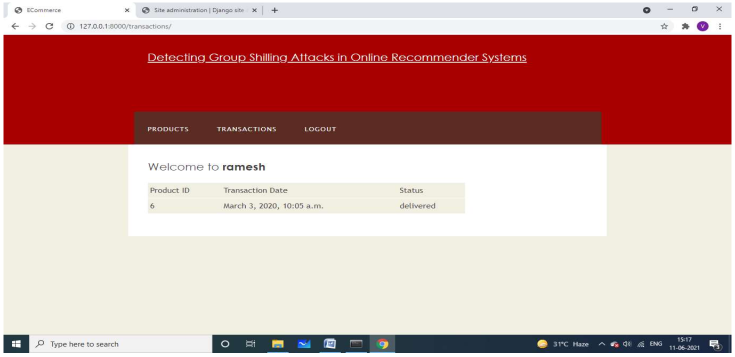


Status :

Image :

Product Name :	oppo mobile
Product Price :	5000
Manufacture :	oppo
Description :	Dazzling White, 4GB RAM, 64GB Storage
Category :	electronics
Name :	
Card NO :	
Cvv :	

Screenshot 6.9 User Buy Product



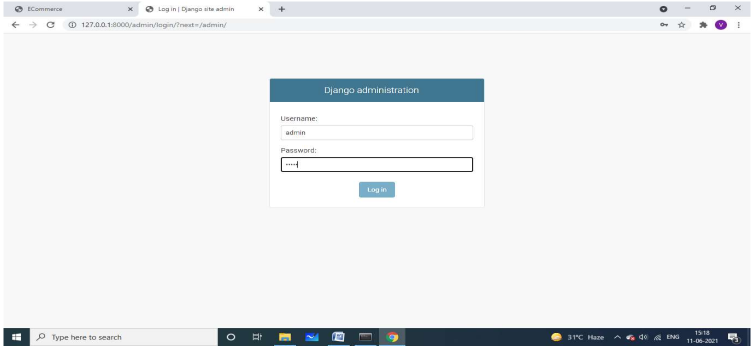
127.0.0.1:8000/transactions/

Detecting Group Shilling Attacks in Online Recommender Systems

Welcome to **ramesh**

Product ID	Transaction Date	Status
6	March 3, 2020, 10:05 a.m.	delivered

Screenshot 6.10 User View Transactions



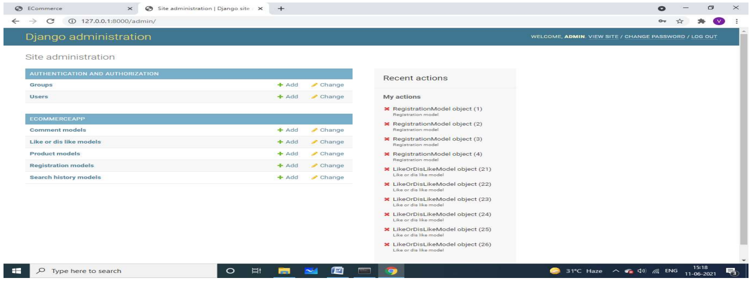
127.0.0.1:8000/admin/login/?next=/admin/

Django administration

Username: admin

Password:

Screenshot 6.11 Database Admin Login



127.0.0.1:8000/admin/

Django administration

WELCOME, **ADMIN**. VIEW SITE / CHANGE PASSWORD / LOG OUT

Site administration

AUTHENTICATION AND AUTHORIZATION		
Groups	+ Add	Change
Users	+ Add	Change

ECOMMERCEAPP		
Comment models	+ Add	Change
Like or dis like models	+ Add	Change
Product models	+ Add	Change
Registration models	+ Add	Change
Search history models	+ Add	Change

Recent actions

My actions

- ✖ RegistrationModel object (1) Registration model
- ✖ RegistrationModel object (2) Registration model
- ✖ RegistrationModel object (3) Registration model
- ✖ RegistrationModel object (4) Registration model
- ✖ LikeOrDisLikeModel object (21) Like or dis like model
- ✖ LikeOrDisLikeModel object (22) Like or dis like model
- ✖ LikeOrDisLikeModel object (23) Like or dis like model
- ✖ LikeOrDisLikeModel object (24) Like or dis like model
- ✖ LikeOrDisLikeModel object (25) Like or dis like model
- ✖ LikeOrDisLikeModel object (26) Like or dis like model

Screenshot 6.12 Database Tables**7.**

Conclusion

Group shilling attacks are a great threat to recommender systems. To detect such attacks, we propose a group attack detection model based on the bisecting K-means algorithm. The proposed detection model can overcome the problem that the performance is poor when attackers have a few corated items. In order to divide candidate groups, we use the fixed time length and dynamically select the starting time point to divide each item's rating track. We combine the features of items and users to calculate the GSDs. Based on the GSDs, the bisecting K-means algorithm is utilized to identify attack groups from the candidate groups. The experimental results on two data sets illustrate the effectiveness of our method.

REFERENCES

I. W. Zhou, J. Wen, Q. Xiong, M. Gao, and J. Zeng, "SVM-TIA a shilling attack detection method based on SVM and target item analysis in recommender systems," Neurocomputing, vol. 210, pp. 197–205, Oct. 2016.

II. W. Li, M. Gao, H. Li, Q. Xiong, J. Wen, and B. Ling, "An shilling attack detection algorithm based on popularity degree features," (in Chinese) Acta Automatica Sinica, vol. 41, no. 9, pp. 1563–1575, Sep. 2015.

III. I. Gunes, C. Kaleli, A. Bilge, and H. Polat, "Shilling attacks against recommender systems: A comprehensive survey," Artif. Intell. Rev., vol. 42, no. 4, pp. 767–799, Dec. 2014.

IV. T. L. Ngo-Ye and A. P. Sinha, "Analyzing online review helpfulness using a regressional relief F-Enhanced text mining method," ACM Trans. Manage. Inf. Syst., vol. 3, no. 2, pp. 10:1–10:20, Jul. 2012.

V. D. Jia, C. Zeng, Z. Y. Peng, P. Cheng, Z. M. Yang, and Z. Lu, "A user preference based automatic potential group generation method for social media sharing and recommendation," (in Chinese) Jisuanji Xuebao, vol. 35, no. 11, pp. 2382–2391, Nov. 2012.

VI. Y. Wang, Z. Wu, J. Cao, and C. Fang, "Towards a tricksy group shilling attack model against recommender systems," in Proc. 8th Int. Conf. Adv. Data Min. Appl., Nanjing, China, 2012, pp. 675–688.

VII. K. Murugesan and J. Zhang, "Hybrid bisect K-Means clustering algorithm," in Proc. Int. Conf. Bus. Comput. Global Informatization, Jul. 2011, pp. 216–219.