Spammers Behaviour Analysis Using Machine Learning Algorithm

Packialakshmi M [1] , Pavithra M[2] , Tamilmozhi T

[1,2,3]IV year Student / Department Of Computer Science And Engineering ,
Sree Sowdambika College of Engineering , Aruppukottai ,

Abstract:

An Industrial Mobile Network helps in securing an industrial production and also helping in the normal function of the machines. Two varied types of users are found in any mobile network: Normal users and Spamming users. The users who create links to viruses and unwanted ads stuffs are called as Spammers. of those Spammers against the normal users in a multi-dimensional data is tedious. To address this, we would like to demonstrate a early demonstrated topic Spatial Identification scheme based on Gaussian Mixture Model (SIGMM). We use the mobile data set obtained from Kaggle network. We use this mobile data set in the SIGMM model for training and test. 70% of the model data is used for training whereas 30% of the data is used for the testing. We then compare the results obtained from SIGMM model with two other newly created models like Reality mining model and hybrid Fuzzy C-Mean model. Comparing down the results prove us that SIGMM model outstands both of those provided models in Recall, Precision and Time complexity.

I. MODULE DESCRIPTION

A. *Project Manager*

this module, project manager only register team leader and team members and they can create group, each group have own team leader. Manager can add team members for each group. Manager can view all files who are uploaded from team leader. Finally he/she only know who is the spammer in our organization.

B. *Team Leader*

In this module, team leader can login to the site through password. They can view group details and upload the file details. Leader can view all file request from team members. If that members are valid leader can send the file key that person. Suppose if the leader can send the file key to other group members they are consider as spammer.

C. *Team Member*

In this module, team member can login to the site through password. They can view group details and file details. And send the file key request to team leader. If the leader send the file key member can download the file through file key. Suppose if the member can send the file key to other group members they are also consider as spammer.

D. *Find Spammer*

In this module, if the team leader or team member can send the file key to other group they are consider as spammer. Project manager only view the spammer details.

E. *Download File*

In this module, team member can download the file through key, that is sended by team leader of their group.

II. EXISTING SYSTEM

The mobile network becomes a target of spammers due to its importance in industrial production control. Spam is one of the most common forms of attack in mobile networks. Spammers pretend to be normal users and only send spam, and these are the users we aim to detect. A serious problem caused by spam is that links leading to viruses are selected by mistake and then users' personal information is stolen, or production control is interfered with. These malicious nodes communicate with each other and spammers hide in them.

Disadvantages of Existing System

1) Classification based on machine learning is a learning process for mapping data samples into two classes. However it has limitations. One is data imbalance, unlabeled data are present in a much larger amount than labeled data.

2) Another limitation is multidimensional data, too many features can lead to overfittinig.

III. PROPOSED SYSTEM

It provides intelligent identification of spammers without relying on flexible and unreliable relationships. SIGMM combines the presentation of data, where each user node is classified into one class in the construction process of the model. We validate

SIGMM by comparing it with the reality mining algorithm and hybrid FCM clustering algorithm using a mobile network dataset from a cloud server.

Advantages of Proposed System

The two other models in terms of identifying spammers and reducing time complexity.

IV. SYSTEM SPECIFICATION

A. Hardware Configuration

The Below Hardware Specifications were used in both Server and Client machines when developing.

Processor	:	Intel(R) Core(TM) i3
Processor Speed	:	3.06 GHz
RAM	:	2 GB
Hard Disk Drive	:	250 GB
CD-ROM Drive	:	Sony
Monitor	:	“17” inches
Keyboard	:	TVS Gold
Mouse	:	Logitech

B. Software Configuration

The Below Software Specifications were used in machines when developing.

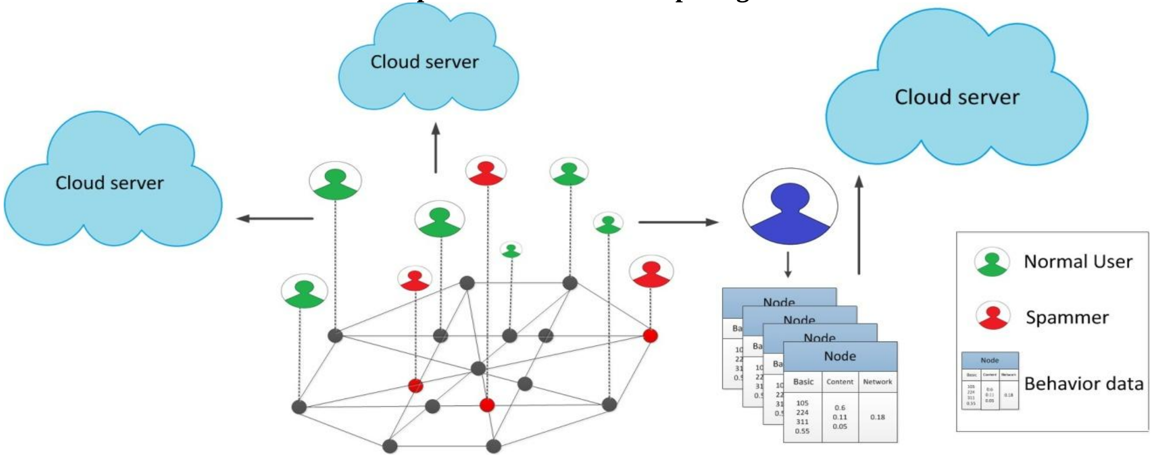
Server

Operating System	:	Windows 7
Technology Used	:	PHP
Database	:	My-Sql
Database Connectivity	:	Native Connectivity
Web Server	:	Apache
Browser	:	Internet Explorer 6.0

Client

Operating System	:	Windows 7
Browser	:	Internet Explorer 6.0

Spammers in cloud computing



REFERENCE

[1] J. Miranda, N. Makitalo, J. Garcia-Alonso, J. Berrocal, T. Mikkonen, C. Canal, and J. M. Murillo, “From the internet of things to the internet of people,” IEEE Internet Computing, vol. 19, no. 2, pp. 40–47, 2015.

[2] T. Qiu, A. Zhao, F. Xia, W. Si, and D. O. Wu, “Rose: Robustness strategy for scale-free wireless sensor networks,” IEEE/ACM Transactions on Networking, vol. 25, no. 5, pp. 2944–2959, 2017.

[3] L. Yao, Q. Z. Sheng, and S. Dustdar, “Web-based management of the internet of things,” IEEE Internet Computing, vol. 19, no. 4, pp. 60–67, 2015.

[4] T. Qiu, R. Qiao, and D. O. Wu, “Eabs: An event-aware backpressure scheduling scheme for emergency internet of things,” IEEE Transactions on Mobile Computing,vol. 17, no. 1, pp. 72–84, 2017.

[5] T. Qiu, K. Zheng, H. Song, M. Han, and B. Kantarci, “A local-optimization emergency scheduling scheme with self-recovery for smart grid,” IEEE Transactions on Industrial Informatics, vol. 13, no. 6, pp. 3195–3205,2017.

[6] S. Lu, V. H. Nascimento, J. Sun, and Z. Wang, “Sparsityaware adaptive link combination approach over distributed networks,” Electronics Letters, vol. 50, no. 18, pP. 1285–1287, 2014.

[7] E. Tan, L. Guo, S. Chen, X. Zhang, and Y. Zhao, “Spammer behavior analysis and detection in user generated content on social networks,” in IEEE International Conference on Distributed Computing Systems, May. 16-18, 2012, pp. 305–314.

[8] P. Heymann, G. Koutrika, and H. Garcia-Molina, “Fighting spam on social web sites,” IEEE Internet Computing, vol. 11, no. 6, pp. 36–45, 2007.

[9] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, “Link prediction using supervised learning,” in Proc of Sdm Workshop on Link Analysis Counterterrorism and Security, Apr. 26-28, 2006, pp. 798–805.

[10] F. Pedregosa, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, “Scikit-learn: Machine learning in python,” Journal of Machine Learning Research, vol. 12, no. 10, pp. 2825–2830, 2013