A STUDY OF DETECTING THE FAKE USERS BY USING UNSUPERVISED LEARNING ALGORITHM

M. Mohan Prasad[1], K. Paneerselvam[2], Dr. N. Revathy[3,] Maria Savari Antony James[4], R. Naveen Kumar[5]
[1, 2, 4, 5] III MCA, [3]Associate Professor
PG and Research Department of Master of Computer Applications,
Hindusthan College of Arts and Science
Coimbatore, Tamilnadu, India

Abstract: In Online Social Network like facebook, twitter has been exposed to more fake account for normal peoples and especially for celebrities. These activities lead to more cyber crime and memory management issues. In order to prevent these kinds of attack, new mechanism has to be proposed to prevent the legitimate user from the malicious user. In this paper, a novel mechanism named as Trust Computation using unsupervised learning model. Fundamental process of this application is to generate reputation of the account through trust computation. The trust computation is enabled with verifier through multiple voting and rating technique on several characteristics such as validating the email address of the user, mobile phone number, schooling details and employment details with friend list and mutual friends in the group. The model effectively isolates fake follower accounts from genuine users. The experimental results prove that proposed system outperforms the existing approaches in terms of accuracy and efficiency parameters.

Keywords: Fake user Detection, Unsupervised Learning.

1. INTRODUCTION

A social networking site is a website where each user has a profile and can keep in contact with friends, share their updates, meet new people who have the same interests. These Online Social Networks (OSN) uses web technology, which allows users to interact with each other. These social networking sites are growing rapidly and changing the way people keep in contact with each other [1]. The online communities bring people with same interests together which makes users easier to make new friends.

In the present generation, the social life of everyone has become associated with the online social networks. These sites have made a drastic change in the way we pursue our social life. Making friends and keeping in contact with them and their updates has become easier. But with their rapid growth, many problems like fake profiles, online impersonation have also grown. There are no feasible solution exist to control these problems. In this project, we came up with a framework with which automatic detection of fake profiles is possible and is efficient. This framework uses classification techniques like Support Vector Machine, Nave Bayes and Decision trees to classify the profiles into fake or genuine classes. The automatic detection method, it can be applied easily by online social networks which has millions of profile whose profiles cannot be examined manually [2].

Trust Computation using unsupervised learning model uses the Fundamental process of this account for trust computation [3]. The trust computation is enabled with verifier through multiple voting and rating technique on several characteristics such as validating the email address of the user, mobile phone no , schooling details and employment details with friend list and mutual friends in the group[4].

The rest of paper is organized into following part; section 2 describes the related work whereas section 3 defines the proposed model. Section 4 evaluates the results and finally section 5 provides the conclusion of the work.

2. RELATED WORK

There exist many literature related to the proposed objective on basis on various learning model which is as follows

2.1. Sybil Account Detection in the Networking

Sybil detectors in social networks leverage the assumption that Sybils will find it hard to befriend real users which leads to Sybils being connected to each other forming strongly connected sub graphs that can be detected using graph theory [5]. The graph theory detects the fake user based on the decision rules.

2.2 Fake Account detection using pattern matching algorithms

Criteria - spam based patterns is used to measure the similarity across accounts in terms of similar name and similar other properties of the account. It is to define the correlation among accounts by investigating their contents. The accounts have similar behaviour if controlled by the same master [6].

3. PROPOSED MODEL

The Proposed model describes the flow of the attack detection with illustration of the model in figure 3.1. Once the Contact is established based response, user will be allowed to share their interest or information mutually

3.1 Profile Creation or Membership Creation

In this Module, User Interface is created to obtain the personnel information of the user to the social network in order to create an account to share their view and information regarding any Business, politics, Climate, Medical Subjects, Social subject or Technical subject in the online environment. This process extracts personnel information such as date of birth, Gender Marital status, College name, School Name etc.

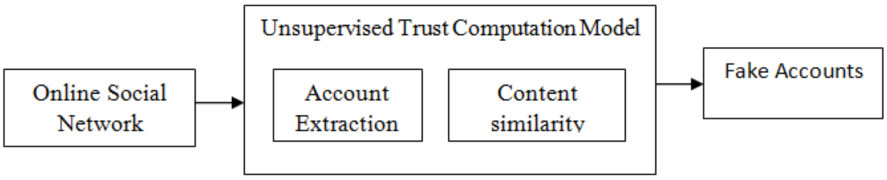

Figure 3.1: Architecture Diagram of the Proposed Model

3.2 Group Establishment or Friend List Creation

The Group establishing is used to set up the medium for information sharing and obtaining the updates of the user in that particular group. The group establishment is carried out using request and response method. Request can established using searching mechanism, Search mechanism is enabled to find the friend or group based on the personnel interest and predefined perspective. Search input is given in terms of Name to be searched, School to be searched, College Name to be searched, City to be searched and Employer to be searched. Based on the imputation, similar result will be retuned based on similarity, request will be send to join the friend list.

3.3 Fake Profile Detection

Fake profiles are the profiles which are not genuine i.e. they are profiles of persons or employer or business who claim to be someone they are not, doing some malicious and undesirable activity, causing problems to the social network and fellow users in order to obtain the information.

4. EXPERIMENTAL ANALYSIS

The experimental evaluation is carried out in several performance measures such as precision and recall and its values have been calculated and tabulated in the table 4.1

Technique	Precision	Recall
Proposed – Trust Computation	98.37	98.56
Existing – Pattern Matching	87.25	88.56

Response to the requested friend is also carried by analysing the similarity between the requester.

5. CONCLUSION

We designed and implemented the fake account detection model using data mining methods. It can detect fake profiles in any online social network with a very high efficiency as high as around 95%. Fake profile detection can be improved by applying most techniques to process the posts and the profile. The trust computation is enabled with verifier through multiple voting and rating technique on several characteristics.

REFERENCE

[1]. M. K. Aguilera, R. E. Strom, D. C. Sturman, M. Astley, and T. D. Chandra, “Matching events in a content-based subscription system,” in Proc. 18th Annu. ACM Symp. Principles Distrib. Comput., 1999, pp. 53–61.

[2]. M. Altinel and M. J. Franklin, “Efficient filtering of XML documents for selective dissemination of information,” in Proc. 26th Int. Conf. Very Large Data Bases, 2000, pp. 53–64.

[3]. B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, “Models and issues in data stream systems,” in Proc. 21st ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst., 2002, pp. 1–16.

[4]. X. Cao, G. Cong, and C. S. Jensen, “Retrieving top-k prestigebased relevant spatial web objects,” Proc. VLDB Endowment, vol. 3, no. 1, pp. 373–384, 2010

[5]. X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi, “Collective spatial keyword querying,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 373–384.

[6]. X. Chen, Y. Chen, and F. Rao, “An efficient spatial publish/ subscribe system for intelligent location-based services,” in Proc. 2nd Int. Workshop Distrib. Event-Based Syst., 2003, pp. 1–6.

[7]. Y.-Y. Chen, T. Suel, and A. Markowetz, “Efficient query processing in geographic web search engines,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2006, pp. 277–288.

[8]. G. Cong, C. S. Jensen, and D. Wu, “Efficient retrieval of the top-k most relevant spatial web objects,” Proc. VLDB, vol. 2, no. 1, pp. 337–348, 2009.

[9]. P. Costa and G. P. Picco, “Semi-probabilistic content-based publish-subscribe,” in Proc. 25th IEEE Int. Conf. Distrib. Comput. Syst., 2005, pp. 575–585.

[10]. G. Cugola and J. E. M. de Cote, “On introducing location awareness in publish-subscribe middleware,” in Proc. IEEE Int. Conf. Distrib. Comput. Syst. Workshops, 2005, pp. 377–382