Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 12, December 2014, pg.151 – 155

RESEARCH ARTICLE

Sanitization Techniques against Personal Information Inference Attack on Social Network

Jyoti B.Mhaske
Computer Department, MET BKC Adgaon, Nashik, Savitribai Phule Pune University, Maharashtra India
jyotibmhaske24@gmail.com

Abstract- Online social networking has become one of the most popular activities on the web. Online social networks such as Facebook are increasingly utilized by many people. OSNs allow users to control and customize what personal information is available to other users. These networks allow users to publish details about themselves and to connect to their friends. Some of the information revealed inside these networks is meant to be private. A privacy breach occurs when sensitive information about the user the information that an individual wants to keep from public is disclosed to an adversary. Yet it is possible to use learning algorithms on released data to predict private information. Private information leakage could be an important issue in some cases. Here the goal is simulate the inference attacks using released social networking data to predict private information. In the proposed system desired use of data and individual privacy presents an opportunity for privacy preserving social network data mining. Here in the system there are three possible sanitization techniques that could be used in various situations for preventing inference attack, those techniques are removing details, adding some new information and manipulate some field these techniques are used for preventing inference attack.

Keywords: Social network analysis, data mining, social network privacy

I. Introduction

Social networking used to connect and share information with friends. People may use social networking services for different reasons to network with new contacts, reconnect with former friends, maintain current relationships, build or promote a business or project, participate in discussions about a certain topic, or just have fun meeting and interacting with other users. Social Network and Twitter have a broad range of users. LinkedIn has positioned itself as a professional networking site profiles include resume information and groups are created to share questions and ideas with peers in similar fields. Unlike traditional personal homepages people in these societies publish not only their personal attributes, but also their relationships with friends. It may causes the privacy violation in social networks. Information privacy is needed for users. Existing techniques are used to prevent direct disclosure of sensitive personal information. Here the focuses on social network data

classification and inferring the individuals private information. More private information are inferred by applying collective classification algorithm.

The system enhance how the online social network data could be used to predict some individual private trait that a user is not willing to disclose (e.g. gender identification, sexual orientation).For example in an office people connect to each other because of similar professions. Therefore it is possible that one may be able to infer someone's attribute from the attributes of his/her friends. In such cases, privacy is indirectly disclosed by their social relations rather than from the owner directly. This is called personal information leakage from inference.

Purpose- Here it has been proposed to design a system that explore the effect of possible data sanitization approaches on preventing such private information leakage, while allowing the recipient of the sanitized data to do inference on non-private details.

Motivation of the Proposed Approach

Desired use of data and individual privacy presents an opportunity for privacy preserving social network. That is the discovery of information and relationships from social network data without violating privacy .But the problem of sanitizing a social network to prevent inference of social network data and then examines the effectiveness of those approaches on a user profile data set. In order to protect privacy, sanitize both details and the underlying link structure of the graph. That is delete some information from a users profile and remove some links between friends. In many situations the data needs to be published and shared with others. Social networks are online applications that allow their users to connect by means of various linktypes. As part of their professional network because of users specify details which are related to their professional life. These sites gather extensive personal information, social network application providers have a rare opportunity direct use of this information could be useful to advertisers for direct marketing. In such situation need to prevent inference attack by using sanitization technique ,sanitize the data set before release to third party ,it is very helpful to user for publishing their details on social site.

II. Literature Survey

Zheleva and Getoor propose several methods of social graph anonymization, focusing mainly on the idea that by anonymizing both the nodes in the group and the link structure, that one thereby anonymizes the graph as a whole. However, the methods all focus on anonymity in the structure itself. For example, through the use of k-anonymity or t-closeness, depending on the quasi identifiers which are chosen, much of the uniqueness in the data may be lost. Through the method of anonymity preservation, maintain the full uniqueness in each node, which allows more information in the data post release[9]

J. He, W. Chu, and V. Liu, the authors conduct distributed data mining in a peer-to-peer network to end usage data about the network itself. However, the situation mentioned in this system is different from 7 scenario; in research the assumption is that the data is distributed fully across the system with each site having only minuscule knowledge of the entirety. The system is take overall social networks data and divided among several data warehouses where perform classification[4].

Gross et al. examine specific usage instances at Carnegie Mellon. Note potential attacks, such as node reidentification or stalking, that easily accessible data on Facebook could assist with. and further note that while privacy controls may exist on the users end

of the social networking site, many individuals do not take advantage of this tool. This finding coincides very well with the amount of data that able to crawl using a very simple crawler on a Facebook network. However, need to extend on work by experimentally examining the accuracy of some types of the demographic reidentification that propose before and after sanitization[10].

III. Proposed System

The system architecture of sanitization method of user profile information is show the detailed work flow of system is as shown in figure4.4 ,how could be gather information of users profile. In that it is important to note that for any detail type, the expected response can either be single or multivalued, and that a user has the option of listing no detail values for any given detail. For example, consider user profile data as home town and activities detail type. A user can only have one home town, but can list multiple activities (for instance, soccer, reading, video games). However, a user also has the option of listing no detail values for these. For example the detail value of video games for the detail type activities will be listed as (activities, video games), to distinguish

it from other details that may have the same detail value, such as (groups, video games). Further even if a user lists multiple activities, data can store each independently in a detail with the corresponding detail name. That is a user who enters jogging and swimming as his favorite activities will have the corresponding details favorite activity, jogging and favorite activity, swimming.

The Naive Bays classifier classify the user profile data into specified classes by using Nave Bayes classifier, Nave Bayes classifier improves the accuracy of classification. Classification is occurs on the basis of node and detail information of user profile data. Details can be manipulated in three ways: adding details to nodes, modifying existing details and removing details from nodes. However here details or link information can broadly classify three methods of classifier into two categories: perturbation and anonymization. Adding and modifying details can both be considered methods of perturbation-that is introducing various types of ”noise” into D to decrease classification accuracies. Removing node however it can be considered an anonymization method. After removal of sensitive information it provide the sanitized dataset for further encryption technique which convert the sanitize dataset into encrypted format for privacy concern.

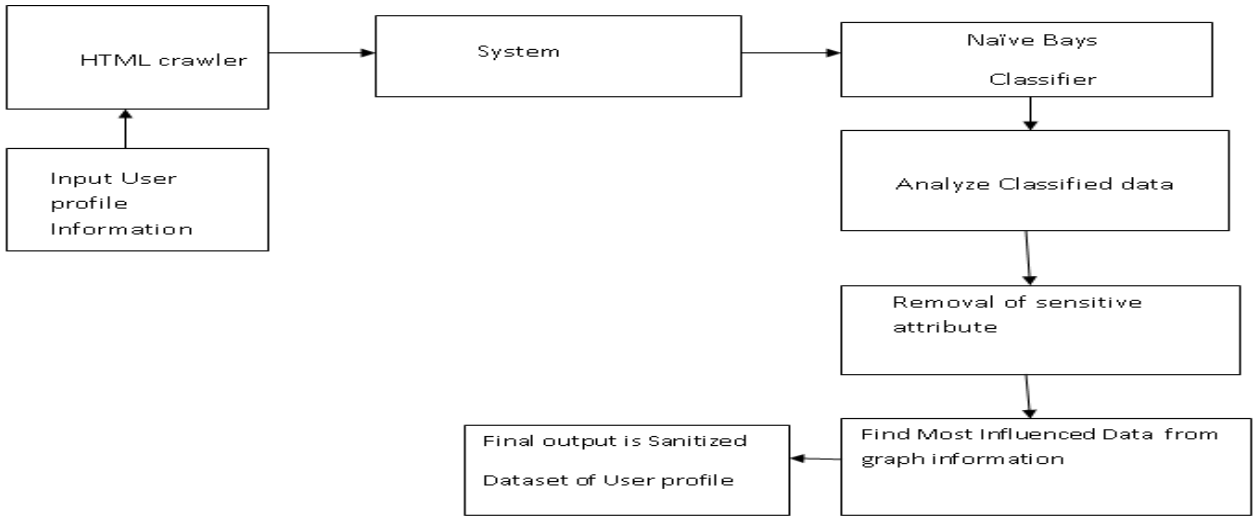

Figure:1 Block Daigram of System architecture

In system architecture First step gather information of users profile. In that it is important to note that for any detail type, the expected response can either be single or multivalued, and that a user has the option of listing no detail values for any given detail. For example, consider Facebook’s “home town” and “activities” detail type. A user can only have one hometown, but can list multiple activities (for instance, soccer, reading, video games). However, a user also has the option of listing no detail values for these. For example, the detail value of “video games” for the detail type “activities” will be listed as (activities, video games), to distinguish it from other details that may have the same detail value, such as(groups, video games). Further, even if a user lists multiple activities, we store each independently in a detail with the corresponding detail name. That is, a user who enters “jogging” and “swimming” as his favorite activities will have the corresponding details (favourite activity, jogging)and (favourite activity, swimming).

The following classifier use for classification purpose:

1.Non-relational (“local”) model. This component consists of a (learned)model, which uses only local information—namely information about (attributes of) the entities whose target variable is to be estimated. The local models can be used to generate priors that comprise the initial state for the relational learning and collective inference components. They also can be used as one source of evidence during collective inference. These models typically are produced by traditional machine learning methods.

2.Relational model. In contrast to the non-relational component, the relational model makes use of the relations in the network as well as the values of attributes of related entities, possibly through long chains of relations. Relational
models also may use local attributes of the entities.

3.Collectiveinferencing. The collective inferencing component determine show the unknown values are estimated together, possibly influencing each other, as described above. Collective inference is a method of classifying social network data using a combination of node details and connecting links in the social graph. Each of these classifiers consists of three components: a local classifier, a relational classifier, and a collective inference algorithm.

For representation work of Social Network the graph algorithm is use which describe status of information of user's profile. Graph representation is very useful for further calculation as link and Details calculation.

The construct graph algorithm:
Input: A degree sequence d of length n.
Output: A graph G(V,E) with node having degree
Sequence d or "No" if the input sequence is not realizable.
1: $V \leftarrow \{1,.....,n\}$, $E \leftarrow \Phi$
2:if $\sum_i d(i)$ is odd then
3: halt and return "no"
4: while 1 do
5: if there exist d(i) such that $d(i)<0$ then
6: halt and return "no"
7: if the sequence d are all zeros then
8: halt and return G(V,E)
9: Pick a random node v with $d(v)>0$
10: set $d(v)=0$
11: $V_{d(v)} \leftarrow$ the d(v)- highest entries in d (other than v)
12: for each node $w \varepsilon V_{d(v)}$ do
13: $E \leftarrow E \cup (v,w) -1$
14: $d(w) \leftarrow d(w)$

IV. Modules

1. Gather input - In this module, collect the data from different users of social network. The crawler loaded all friends of the current profile and stored the friends inside the database both as friendship links and as possible profiles to later crawl.

2. Classification Module- In this module, all information of facebook users classify by using Naïve Bayes classifier on the basis of Links details and Detail type of attribute.

3.Finding sensitive information–In this module find the very sensitive information which can find political affiliation & sexual orientation of users and that information is used for removal module.

4. Removal of detail type and Removal of Link- The other option for anonymizing social networks is altering links. Unlike details, there are only two methods of altering the link structure: adding or removing links.

V. Methodology

In this experiments, we implemented four algorithms to predict the political affiliation of each user. The first algorithm is called "Details Only." This algorithm uses Naïve bayes classifier to predict political affiliation and ignores friendship links. The second algorithm is called "Links Only." This algorithm uses Link based information to predict political affiliation using friendship links and does not consider the details of a person. The third algorithm is called "Average." Average algorithm take the average Details Only and Links Only algorithms, respectively. The final algorithm is a traditional naıve Bayes classifier, which we used as a basis of comparison for our proposed algorithms

VI. Conclusion and future Scope

The System is use for preventing inference attack on user profile data of social network. The proposed system is using both friendship links and details together gives better predictability than details alone. In addition implement the effect of removing details and links in preventing sensitive information leakage. Here discovered situations in which collective inference does not improve on using a simple local classification method to identify nodes. When combine the results from the collective inference implications with the individual results, begin to see that removing details and friendship links together is the best way to reduce classifier accuracy. Here show that each of these methods provides a measure of privacy guarantee for users within the network, but can also be used by third parties for classification on non-sensitive attributes. In future the system can extended for sanitization of prevent private information inference attack is by providing user profile information to third party in encrypted format for maintaining privacy for user profile data.

REFERENCES

[1] "Preventing Private Information Inference Attacks on Social Networks Ray- mond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham," *Fellow,IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 8 AUGUST 2013.*

[2] " J. Lindamood, R. Heatherly, M. Kantarcioglu, and B. Thuraising- ham, Inferring Private Information Using Social Network Data, *Proc.18th Intl Conf. World Wide Web (WWW), 2009.*

[3] " E. Zheleva and L. Getoor, Preserving the Privacy of Sensitive Relationships in Graph Data, *Proc.First ACM SIGKDD Intl Conf. Privacy, Security, and Trust in KDD, pp. 153-171, 2008.*

[4] J. He, W. Chu, and V. Liu, Inferring Privacy Information from Social Networks, *Proc. Intelligence and Security Informatics, 2006.*

[5] K.M. Heussner, Gaydar n Facebook: Can Your Friends Reveal Sexual Orienta- tion ABC News, *http://abcnews.go. com/Technology/gaydar-facebook-friends/storyid=8633224. UZ939UqheOs, Sept. 2009.*

[6] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, Anonymizing Social Networks,*Technical Report 07-19, Univ. of Massachusetts Amherst, 2007.*

[7] K. Liu and E. Terzi, Towards Identity Anonymization on Graphs, *Proc. ACM SIGMOD Intl Conf. Management of Data (SIGMOD 08), pp. 93-106,2008.*

[8] J. He, W. Chu, and V. Liu, Inferring Privacy Information from Social Networks, *Proc. Intelligence and Security Informatics, 2006.*

[9] E.Zheleva and L. Getoor, Preserving the Privacy of Sensitive Relationships in Graph Data, *Proc.First ACM SIGKDD Intl Conf. Privacy, Security, and Trust in KDD, pp. 153-171, 2008.*

[10] R. Gross, A. Acquisti, and J.H. Heinz, Information Revelation and Privacy in On- line Social Networks, *Proc.ACM Workshop Privacy in the Electronic Soc. (WPES 05), pp. 71-80, http://dx.doi.org/10.1145/1102199.1102214, 2005.*

[11] J. Yedidia, W. Freeman, and Y. Weiss. Exploring Artificial Intelligence in the New Millennium. Science Technology *Books, 2003.*

