International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 8, August - 2014

Privacy Preserving Data Mining for Social Networks

Ms. Brinal Colaco
M.E. Student in Computer Engineering Dept. Faculty of Computer Engineering
St. Francis Institute of Technology
Mumbai University
Mumbai-400103. India.

Mr. Shamsuddin Khan
Assistant Professor,
Department of Computer Engineering
St.Francis Institute of Technology,
Mumbai University
Mumbai- 400103, India

***Abstract*—Advances in technology has made it possible for hackers and intruders to collect personal and professional data about individuals and the connections between them, such as their email correspondence and friendships on the internet. These hackers could either be third party agencies or individuals who are interested in knowing more about the users of the social networks. Most of the information present in these social networks is not private, yet learning algorithms could be used on the public data to predict private information. This paper focuses on the problem of private information leakage from the information present on the social networks. It represents the cause-effect relationships within social network data by the application of the soft computing technique of fuzzy Inference Systems. Sanitization techniques that could be used in various inference attack scenarios are suggested and effectiveness of these sanitization techniques is analyzed.**

***Keywords*—Social Network Security, Inference attacks, Fuzzy Inference System**

I. INTRODUCTION

Social network analysis is the methodical analysis of social networks. In social networks, the individual entities within the networks are considered as nodes and the relationship between the entities are considered as links between the nodes [1][2].Various details like the hobbies, music interests, movies, books, favorite activities, professional information, etc. of the user is present in the user profile of the individual[3]. Social networks gather extensive personal information because of which the application providers have the opportunity of using the present information. However, in practice, privacy concerns can prevent these efforts [4][5]. The conflict between the desired use of data and individual privacy presents an opportunity for privacy-preserving social network data mining i.e. the discovery of information and relationships from social network data without violating privacy of the individual [3].In this research, the focus is on predicting private information using public information of the user that is present on the social network. In a social network, users have profile data that make certain aspects of their personality predictable. The identification or prediction of the underlying private attributes could have negative repercussions. For example, it is possible to determine a user's sexual orientation by obtaining few details from Facebook like user's gender, the gender they are interested in and the same details of the friends in their Friend list, or a user showing interest in a few groups on the social networks can predict his/her affiliation to a certain political group [6] [7].The inference of private information from the information present on the social networks is a major security breach. For this, a particular class of attacks are explored, namely, the class of Inference Attacks. These attacks are used to gain knowledge about a subject or a database. The attacks make an attempt to deduce sensitive information from trivial information that is publically available. The main goal of this work is to present a method based on the Fuzzy Inference System (FIS) that can be applied for the development of an expert system for predicting private information of an individual and also suggest ways to avoid the inference attack. No previous work has been done on implementation of FIS techniques to handle the specific problem for inference attacks on social networks. Efficiency of this technique will be compared to the efficiency of other techniques of Inference Attacks on Social Networks (e.g. Naïve Bayesian Classification technique)

II. RELATED WORK

The area of privacy within a social network is very vast and covers many research areas. Hay et al. [8] considers several ways of anonymizing social networks but the research mentioned in this paper is related to inferring the details on the social network not individually identifying individuals.Other authorshave tried to infer private information of the social networks. In [8], He et al, consider ways to infer private information via friendship links by creating a Bayesian network from the links inside a social network. Many inference attacks can be performed using third party extensions. In [3], Heatherly et al have workedon

Launching inference attacks using social networking data to predict private information. This paper also discusses sanitizing a social network to prevent inference of social network data. It examines the effectiveness of those approaches on a real-world data set. The authors use Naïve Bayes Classification and other hybrid classification techniques for inference. In [10], Ahmadinejad et al study inference attacks that can be launched via the extension Application Programming Interface (API) of Facebook. Taxonomyof such attacks is devised anda risk metric is proposed to help subscribers of the third party applications refine their privacy expectations. In [11], Lei et al, write about the mutual friend feature based attacks on social networks. The mutual friend feature raises significant privacy concerns as an adversary can use it to find out some or all of the victim's friends, although, as per the privacy settings of the victim, the adversary is not authorized to see his friend list directly. It focuses only on the mutual friends feature where as this paper considers any information including the mutual friends. None of the above mentioned works have made use of the Fuzzy Inference System technique. This paper focuses entirely on the use of Fuzzy Inference System for prediction of private information from the available public information. FIS for decision support was chosen because of the nature of the application and due to the reason that prediction of private information is a complex process with sufficient interacting parameters and Fuzzy Inference systemsare suitable for this kind of problems. FIS also becomes a good choice because of the ease of use and the low time requirement for computation [12][13].

III. PROPOSED WORK

The proposed work in this paper focuses on determining whether the public data present on the social networks is vulnerable enough to predict sensitive private information and thereby suggest ways to avoid the same.
The implementation of the proposed system contains the following core modules:
Gather data from social network: A database of 2000Indian Facebook profiles was made to test on the proposed system. This database consists of basic user
information like the name of the user, the gender, state, groups liked on Facebook , music interest of the
user, pages liked on Facebook, number of friends on Facebook, gender the user is interested in, etc. Three private attribute predictions are made namely religion, political affiliation and sexual orientation. The religion is predicted using available public information like the music interest of the user and the state the user belongsto. Political affiliation is predicted using the religion and the related political groups that are liked on Facebook. Sexual orientation is predicted using the related groups, pages liked and

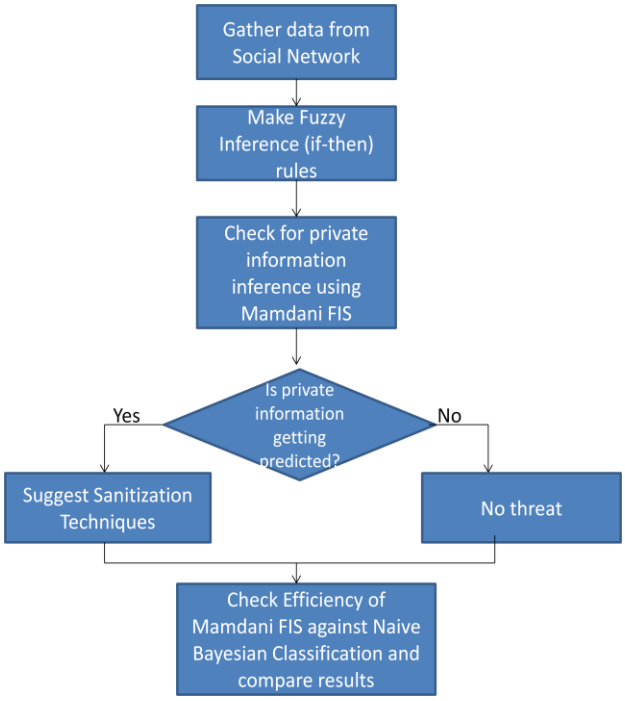

FIGURE 3.1 PROPOSED SYSTEM

the number of asexual/bisexual/ homosexual friends on Facebook.

Make fuzzy if-then rules: Fuzzy inputs are given to the fuzzy inference system and as Mamdani fuzzy inference is used, the output obtained is also fuzzy. In order to obtain the output, fuzzy if-then rules are made. The correct output will be obtained based on these rules. For example,

Table 3.1

Detail Name	Detail group	Music Interest	State	Religion	Political Affiliation
Radhika G.	ShivajiMaharaj Group	Indian Classical Music	Mah	Hindu	Shiv sena

As shown in the table above the religion is the private attribute that could be predicted using the available/public attributes like state and music interest of the user. So, for somebody belonging to the state Maharashtra and having an interest in Indian classical music or Bhajan would belong to the Hindu religion, and for someone with Hindu as their religion and ShivajiMaharaj group as their favourite group would belong to the Shiv Sena political party. All the rules for prediction are made keeping the general scenarios. It may not be true for all the users. Also, the fuzziness of the attributes comes from the overlapping membership functions assigned to it. For example, consider the border between Maharashtra and Goa, mostly the people residing at the border will have influence of both the cultures. The religion in this case, will be predicted as Maharastrian- Catholics.

Check for private information inference using Fuzzy Inference System: After setting up a rule base for all the attributes available, the system computes the private attributes as per the availability of data. As the attributes are linked, Political affiliation cannot be predicted it religion of the user is not present as an input. In many cases, the most limited and the most private attributes are for sexual orientation. But sexual orientation can be predicted using single attributes like 'interested in- men or women'.

Suggest ways for Sanitization: Sanitization means taking appropriate measures to avoid or prevent the inference attacks. These suggestions could be different for different users based on the private information that is inferred or predicted. Most of these suggestions are as simple as making a small change like deleting some information from the user profile of the social network or making some attributes of the user profile private.

Check the efficiency of Fuzzy Inference System against Naïve Bayesian classification: Inference of private attributes is also obtained using the Naïve Bayesian classification which is probability based classification technique. In case of the fuzzy inference system, better accuracy is obtained as compared to the Naïve Bayesian system. Three parameters are computed- accuracy, precision, recall.

IV. EXPERIMENTAL RESULT

This paper focuses on inference attacks on social networks. These inference attacks are carried out using Mamdani fuzzy inference system and Naive Bayesian classification. If the private attributes are inferred correctly, the system will provide with suggestions as per the predictions in order to avoid inference attacks. The performance of both fuzzy inference system and Naive Bayesian classification technique are tested to determine the accuracy, precision and recall [15]. The results for different range of users are show in the tables below.A dataset with profile information of 2000 users was used for the experiments. General details of the Social Network dataset are provided in Table 1.

Table 3.2

Detail Name	Detail group	Gender	Music Interest	State
Radhika G.	ShivajiMaharaj Group	Female	Indian Classical Music	Maharashtra
Sumit N.	Apositive	Male	Indian Rock	Maharashtra

Two algorithms are implemented to predict the private information from the available public information. The dataset implemented on Mamdani FIS[16] has high prediction accuracy. Sexual orientation of an individual is also predicted based on group likes and the number of asexual/bisexual/homosexual friends in the friend list of an individual. Similar prediction is done using the Naïve Bayesian classification. Parameters like accuracy precision and recall of these two techniques are compared based on these predictions.

A dataset of 2000 entries is used for both the approaches- Mamdani and Naive Bayesian. The results are calculated to determine the accuracy, precision and recall.

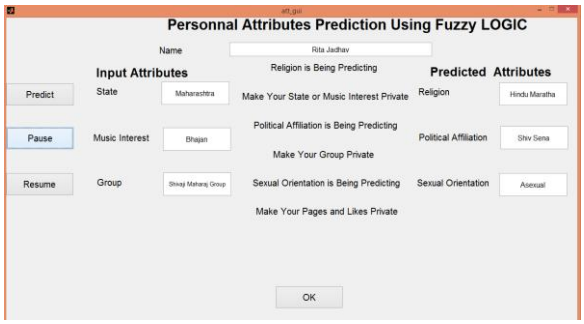

Figure 4.1 Prediction usng Fuzzy Inference system

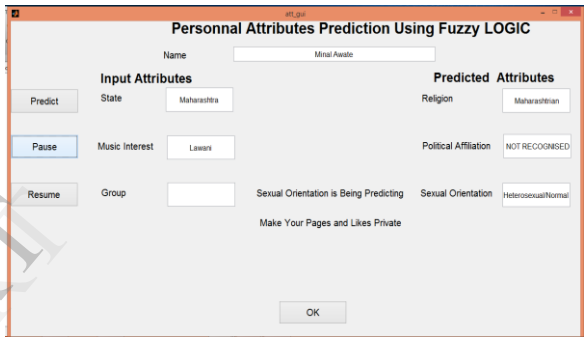

Figure 4.2- Prediction using FIS after sanitization

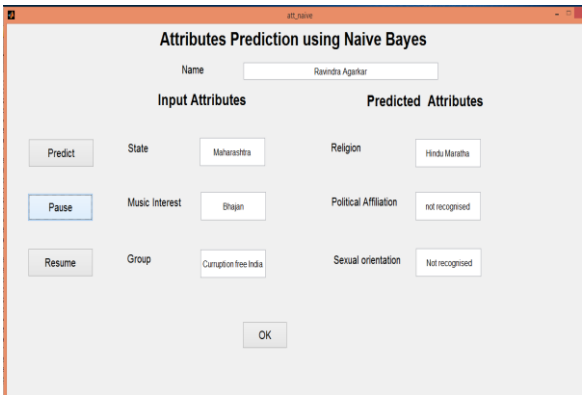

Figure 4.3 Prediction usng Naïve Bayesian Classification

Figure 4.1 shows prediction of private attributes using Mamdani fuzzy inference system along with sanitization suggestions. Based on the input attributes, private attributes can be predicted and the suggestion changes based on the predictions. Figure 4.2 shows the results after sanitization. Figure 4.3 shows prediction using Naïve Bayesian classification technique.

As seen in Table 3.3 , the accuracy of both the techniques is the same.The precision and recall changes for political affiliation and sexual orientation predictions.

Table 3.3: Results for 10 entries **without** Sanitization

	Fuzzy (Mamdani)			**Naive Bayesian**		
	Accuracy	Precision	Recall	Accuracy	Precision	Rec-all
Religion Prediction	100	100	52.9	100	100	50
Political Affiliation	100	90	52.9	90	81.818	56.2
Sexual Orientation	88.89	80	57.14	30	17.64	21.42

Table 3.4 Results for 1000 entries **without** Sanitization

	Fuzzy (Mamdani)			**Naive Bayesian**		
	Accuracy	Precision	Recall	Accuracy	Precision	Rec-all
Religion Prediction	97.79	97.8	50.5	98.19	96.46	50.93
Political Affiliation	97.29	97.39	50.67	64.5	47.6	91.1
Sexual Orientation	7.61	7.60	8.23	5.3	2.72	2.8

In Table 3.4, the accuracy, precision and recall of sexual orientation prediction is very low because in the database, the data for prediction of sexual orientation is limited to 100 users due to the high level of sensitivity. The tables below show the results after making certain attributes private i.e. sanitization.

Table 3.5 and Table 3.6 consists of results post sanitization. As seen in the results, the accuracy and precision of both the techniques after sanitization reduces considerably. The fuzzy inference system gives better results as compared to Naïve bayesian classification even after suggesting sanitization techniques.

Table 3.5 Results of 10 entries with Sanitization (5 political groups deleted/ made private)

	Fuzzy (Mamdani)			**Naive Bayesian**		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Religion Prediction	49.79	49.8	99.4	51.45	34.63	52.98
Political Affiliation	49.29	49.3	97.4	33.7	20.2	25.4
Sexual Orientation	88.89	80	57.14	30	17.64	21.42

Table 3.6: Results for 1000 entries **with** Sanitization (500 states removed)

	Fuzzy (Mamdani)			**Naive Bayesian**		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Religion Prediction	49.79	49.8	99.4	51.45	34.63	52.98
Political Affiliation	49.29	49.3	97.4	33.7	20.2	25.4
Sexual Orientation	-	-	-	-	-	-

V. CONCLUSION

This paper focuses on addressing various issues related to private information leakage in social networks. Inference attacks are one of the most critical attacks that lead to social network data abuse. It is shown how Mamdani FIS is an efficient technique to predict more accurately leading to private information leakage. This technique is compared with the Naïve Bayesian technique on the same dataset. And from the results obtained, it is evident that, before and after sanitization, the fuzzy inference system gives better accuracy and precision. The proposed system determines whether the user profile is vulnerable to inference attacks.Simultaneously, the system gives suggestionsfor hiding the information that lead to the prediction of private attributes. This system will decrease the accuracy of the classifiers that infer private information. This system will try to reduce the inference attacks on social networks, thus enhancing the experience of the user. Further work can be done to tackle the cold start problem wherein a user's profile could be vulnerable to inferences attacks even without sufficient information.

REFERENCES

[1] Pinheiro, Carlos A.R. (2011). Social Network Analysis in Telecommunications. John Wiley & Sons. p. 4. ISBN 978-1-118 01094-5.
[2] D'Andrea, Alessia et al. (2009). "An Overview of Methods for Virtual Social Network Analysis". In Abraham, Ajith et al. Computational Social Network Analysis: Trends, Tools and Research Advances. Springer. p. 8. ISBN 978-1-84882-228-3.
[3] Raymond Heatherly, Murat Kantarcioglu, and BhavaniThuraisingham, "Preventing Private Information Inference Attacks on Social Networks", Knowledge and Data Engineering, IEEE Transactions on Volume: 25, Issue: 8 Publication Year: 2013 , Page(s): 1849 – 1862
[4] PoojaShelke, AshishBadiye, "Social Networking: Its Uses and Abuses", Research Journal of Forensic Sciences, Nagpur, Maharashtra, (2013): 2-7.
[5] Facebook Beacon, 2007.
[6] C. Johnson, "Project Gaydar," The Boston Globe, Sept. 2009.
[7] K.M. Heussner, "'Gaydar' n Facebook: Can Your Friends Reveal Sexual Orientation?" ABCNews, http://abcnews.go.com/Technology/gaydar-facebook-friends/story?id=8633224#.UZ939UqheOs, Sept. 2009.
[8] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, "Anonymizing Social Networks," Technical Report 07-19, Univ. of Massachusetts Amherst, 2007.
[9] He, Jianming, Wesley W. Chu, and Zhenyu Victor Liu. "Inferring privacy information from social networks." Intelligence and Security Informatics. Springer Berlin Heidelberg, 2006. 154-165.
[10] Ahmadinejad, SeyedHossein, and Philip WL Fong. "On the feasibility of inference attacks by third-party etensions to social network systems." Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security. ACM, 2013.
[11] Jin, Lei, James Joshi, and Mohd Anwar. "Mutual-friend Based Attacks in Social Network Systems." Computers & Security (2013).
[12] Jang, Jyh-Shing Roger, Chuen-Tsai Sun, and EijiMizutani. "Neuro fuzzy and soft computing- a computational approach to learning and machine intelligence [Book Review]." Automatic Control, IEEE Transactions on 42.10 (1997): 1482-1484.
[13] Chai, Yuanyuan, LiminJia, and Zundong Zhang. "Mamdani based model adaptive neural fuzzy inference system and its application." International Journal of Computational Intelligence 5.1 (2009): 22-29.
[14] Yager, Ronald R., Dimitar P. Filev, and Tom Sadeghi. "Analysis of flexible structured fuzzy logic controllers." Systems, Man and Cybernetics, IEEE Transactions on 24.7 (1994): 1035-1043.
[15] "Evaluating a Classification Model", http://www.cs.odu.edu/~mukka/cs495s13/Lecturenotes/Chapter5/recallprecision.pdf, 12th May 2014.
[16] Jang, Jyh-Shing Roger, and Chuen-Tsai Sun. Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice-Hall, Inc., 1996.

BrinalColaco *received her B.E. in May 2010 in the field of Information technology from the University of Mumbai. Currently she is pursuing M.E. degree from the University of Mumbai and is also working as a lecturer in St. Francis Institute of Technology.*

Shamsuddin S. Khan *is currently assistant professor at St. Francis Institute of Technology Mumbai in computer engineering dept. His areas of interests include artificial intelligence, neural network, database systems, data mining and distributed computing.*