International Journal of Database Theory and Application
Vol.7, No.3 (2014), pp.121-130
http://dx.doi.org/10.14257/ijdta.2014.7.3.12

Eliminating unwanted messages in SNS using Decision tree

Venkata Naresh Mandhala[1], Debnath Bhattacharyya[2], Tai-hoon Kim[3*]

[1]*Information Technology Department, VFSTR University, Vadlamudi-522213, Guntur, India*
[2]*Department of Computer Application,RCC Institute of Information Technology, Kolkata-700015, India*
[3]*Department of Convergence Security, Sungshin Women's University, 249-1, Dongseon-dong 3-ga, Seoul, 136-742, Korea*
[1]*mvnaresh.mca@gmail.com,* [2]*debnathb@gmail.com,* [3]*taihoonn@daum.net*

Abstract

Currently, the usage of social networking sites are increasing rapidly. Social networks are used to connect people, share information between users and maintain relation between friends. Hundreds of thousands of people are using these social networking services like Facebook, twitter etc., User security is the main aspect in present days from person to person interaction. There is a major task of online social network is information filtering. An online social network provides the little support for allowing sharing the information on the user walls. In this paper we propose to introduce efficient user security mechanism in transmission of messages from one user to another user. So we describe to develop methodology called machine learning algorithm called decision tree algorithm for security in their user walls. Our experimental results show efficient learning process in extracting feedback process generation in customized system operations in online social networks. In this we also consider the functionalities of each user in relevant data sharing of social networks.

Keywords: *Online social network, Privacy, User Security, authentication, Decision tree*

1. Introduction

Social networking is a way for the people to connect each other and share information in online. Now a days millions of people accessing the web world widely regularly from mobile devices, web sites *etc.*, [4]. Many people are using social networking services for many reasons. Those are for to connect with new friends or to share important information between friends and maintain relationship and to have fun meeting with other users *etc.*, Some services like Facebook, twitter and linked in have millions of users. For example it is the professional networking website, it includes resume information, share questions and to meet business people *etc.*, More over we have special social networks to meet our former classmates [4].

The growth of social networking is increased over the last year. In particular, the usage of Facebook is increased rapidly and it has wide range of users internationally. In worldwide, mobile users are active in face book users [5]. Many applications have already taken rather simple and traditional approaches to integrating social network information with user location and context information. The most common form of application simply extends access to social networks to mobile phones or provides social network interfaces optimized for access from these mobile phones [5].

* Corresponding Author

ISSN: 2005-4270 IJDTA

1.1. Privacy in a Connected World - Data Mining in Social Networks

Social network has prosperity of personal information. Some of that information would not be valuable by itself but having a clear picture of everything about a person can give attackers ideas and information required to perform other attacks [5]. In addition to this, underground forums sell personal information. Your data can be mined and stored somewhere in the dark corners of the Internet waiting for a criminal to pay the right price for it. Criminals can use this information to obtain birth certificates/passports/other documentation and fake real-life identities. Some countries have looser controls than others, but in general, identity theft is something that already happens regularly. Another factor that exacerbates this massive data-leak - age potential is a user's public profile. When users set their information to be accessible without logging in to the social networking site, that information can be indexed in search engines or any other archive. There are social networking search engines that can search all available data about any name in a certain region. This makes the lives of stalkers, fraudsters, or any other attacker much easier.

1.2. Security

In addition to privacy concerns social networking sites can be used by cyber criminals to attack you or your devices. Here are some protection steps:

1.2.1. Log in: Protect our social networking account with a strong password and do not share this password with anyone. In addition to this some social networking sites are providing strong authentication, such as two step verification.

1.2.2. Encryption: Many social networking sites allow you to use encryption called HTTPS to secure your connection to the site. Some sites like Twitter and Google+ have this enabled by default, while other sites require you to manually enabled HTTPS via account settings.

1.2.3. Apps: Some social networking sites give you the ability to add or install third - party applications, such as games. Keep in mind there is little or no quality control or review of these applications; they may have full access to your account and private information. Only install apps that you need, that are from well known, trusted sites and remove them when you no longer need them.

Social networking sites are a powerful and fun way to communicate with the world. If you follow the tips outlined here, you should be able to enjoy a much safer online experience.

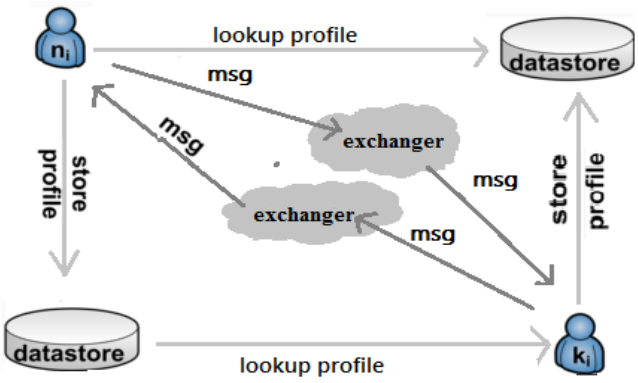

Figure 1. A Secure and Privacy Preserving

2. Previous Works

According to Mindi McDowell and Damon Morda, social networking is a way to connect millions people and share information with each other in online. Thousands of people using these social networking services worldwide rapidly. When you are going to share information between users in social networking sites, you need to follow the potential risks. And you need to beware of what you are going to share [4].

In the view of Aaron Beach, Mike Gartrell, and Richard Han, Social network information is now being used in ways for which it may have not been originally proposed. In specific, increased use of smartphones capable of running applications which access social network information enable applications to be aware of a user's location and preferences. However, current models for exchange of this information require users to compromise their privacy and security [1]. We present several of these privacy and security issues, along with our design and implementation of solutions for these issues [5]. Our work allows location-based services to query local mobile devices for users' social network information, without disclosing user identity or compromising users' privacy and security. We contend that it is important that such solutions be accepted as mobile social networks continue to grow exponentially.

Online social networks are now used by hundreds of millions of people and have become a major platform for communication and interaction between users [2, 4]. This has brought a wealth of information to application developers who develop on top of these networks. Social relation and preference information allows for a unique breed of application that did not previously exist. Furthermore, social network information is now being correlated with users' physical locations, allowing information about users preferences and social relationships to interact in real-time with their physical environment [5].

Giles Hogben, ENISA said, Social networking is becoming the preferred (by end-users) way to manage personal data. It is an area where people take an active interest in how their personal information is managed and displayed rather than being passive account-holders as in most identity management systems. Social engagement provides a much-needed incentive for end-users to engage in processes such as setting privacy rules and providing feedback on spammers [7]. As previously mentioned, social networks represent the world's largest body of personal data.

As Ted Demopoulos said, social networking is a Social networking sites such as Facebook, Twitter, Google+, Pinterest and LinkedIn are powerful, allowing you to meet, interact and

share with people around the world. However, with all these capabilities come risks; not to just you, but your family, friends and employer. In this newsletter we will discuss what these dangers are and how to use these sites more safely.

Content based filtering is an existing system. Where, Information filtering systems are designed to classify a stream of dynamically generated information transmitted asynchronously by an information producer and present to the user those information that are likely to satisfy his/her requirements [3].

In this section we describes the filtered wall architecture with short text classifier users interact with the system by means of a GUI to set up and manage their FRs/BLs. The online social networks extend the process in different situations for accessing services from individuals in their user walls in same situational environment [6, 9].

As shown in Figure 2, filter wall architecture consists following things for efficient accessing in real time applications like face book, and other social networks [10]. First layer in the OSN commonly provides basic functionalities with profile and relationship management and also specifies large number of other network services in external way in real time process generations [11]. The core components of the proposed system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier (STC) modules [8]. The latter component aims to classify messages according to a set of categories. In contrast, the first component exploits the message categorization provided by the STC module to enforce the FRs specified by the user.

3. Proposed System

In this context, many empirical studies have shown that average OSN users have difficulties in understanding also the simple privacy settings provided by today OSNs [12]. To overcome this problem, a promising trend is to exploit data mining techniques to infer the best privacy preferences to suggest to OSN users, on the basis of the available social network data [12].

3.1. Decision Tree

Decision trees are designed essentially for a hierarchical decomposition of the data space. Based on the attribute value it determines the predicate or a condition. In this decision trees, class labels in the leaf node used for classification purpose. In order to reduce the over fitting data, pruning is to be done. There are several different kinds of splits in the decision trees are available [10].

The use of a decision tree is a very popular technique in data mining. In the opinion of many researchers, decision trees are popular due to their simplicity and transparency. Decision trees are self-explanatory; there is no need to be a data mining expert in order to follow a certain decision tree. Classification trees are usually represented graphically as hierarchical structures, making them easier to interpret than other techniques. If the classification tree becomes complicated (*i.e.*, has many nodes) then its straightforward, graphical representation become useless.

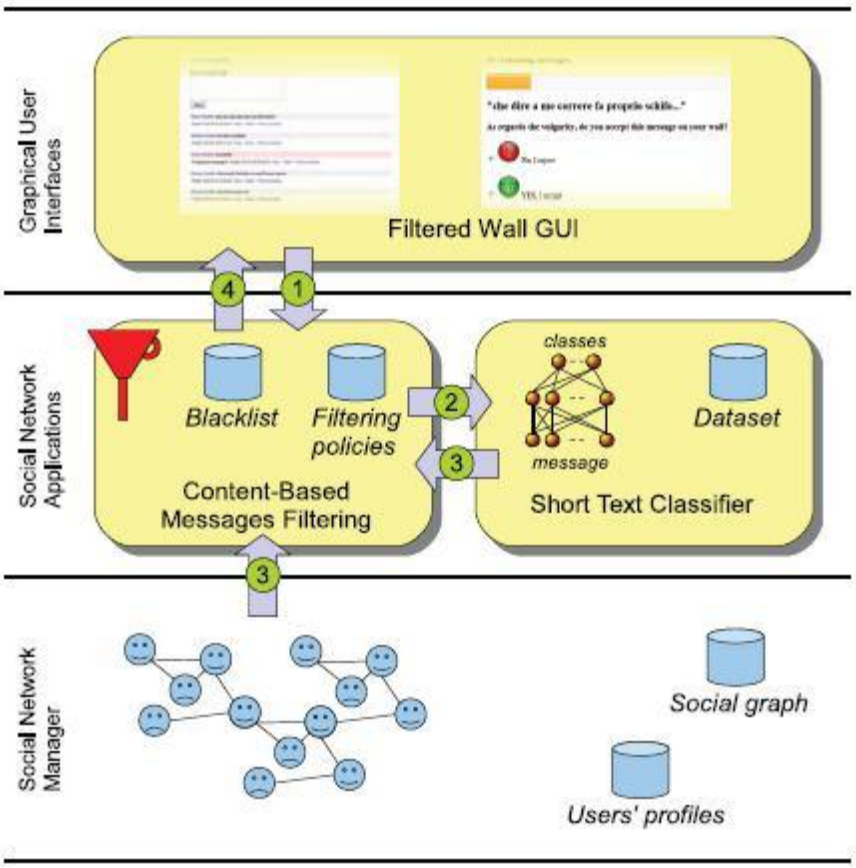

Figure 2. Filtered wall Conceptual Architecture and the Flow Messages

A decision tree represents a procedure for classifying categorical data based on their attributes. It is also efficient for processing large amount of data, so is often used in data mining application. The construction of decision tree does not require any domain knowledge or parameter setting, and therefore appropriate for exploratory knowledge discovery. Their representation of acquired knowledge in tree form is intuitive and easy to assimilate by humans.

3.2. Decision Tree Algorithm

Algorithm: Generate _decision _tree: Generate a decision tree from training tuples of data partition A.

Input:

- Data partition, A, which a set of training tuples and their associated class labels;
- *attribute_list* , the set of candidate attributes:
- *Attribute_selection_method* , a procedure to determine the splitting criterion that best partitions the data tuples into individual classes. This criteion consists of a *splitting_attribute* and, possibly a *split point* or *splitting subset*

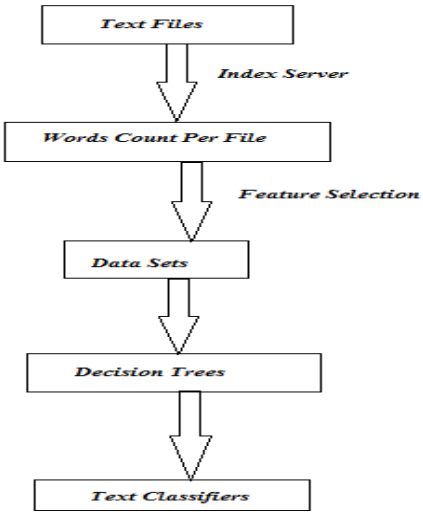

Figure 3. Schematic Learning Process

Method:

(1) Create a node

(2) If tuples in A are all of the same class. B then

(3) *return n* as leaf node labled with the class B;

(4) If attibute list is empty then

(5) return N as leaf node labeld with the majority class in A;// majority voting

(6) apply attribute_selection_method(A,*attribute_list*)to find the "best"*splitting _criterion*;

(7) label node N with *splitting_criterion*;

(8) If *splitting_attribute* is decrete-valued and multi way splits allowed then//not restricted to binary trees.

(9) *attribute_list attribute_list_splitting _attribute*;//remove *splitting_attribute*

(10) For each criterion f of *splitting_criterion*

//partition the tuples and grow the subtress for each partition

(11) Let A_f be the data tuples in A satisfying the out come f;//a partition

(12) If A_f is empty then

(13) Attach a leaf labeled with majority class in A to node N;

(14) else attach the node returned by Generate_decision_tree (A_f_*attribute_list*) to node N;

end for

(15) return N;

4. Result and Analysis

This paper explains about the separation of user unwanted and wanted messages in social networking sites. The generation of decision tree from wanted and unwanted messages in a data partition set A. Attribute_list indicates the set of wanted messages in data partition set A. Attribute_selection_method determines the separation of unwanted messages from data partition set A.

For example consider a G-mail account. In which we get valid and invalid messages. In the above algorithm we consider A as a set of valid messages and B as set of invalid messages. If we get a valid message then it will go to set A by using attribute selection method otherwise it will go to set B [5].

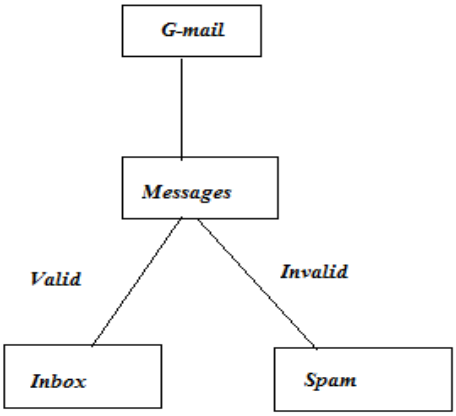

Figure 4. Filtering Unwanted Messages

We can compare the existing system with proposed system given in Figure 5. In Proposed system by using decision trees we can get effective results compared to the existing system.

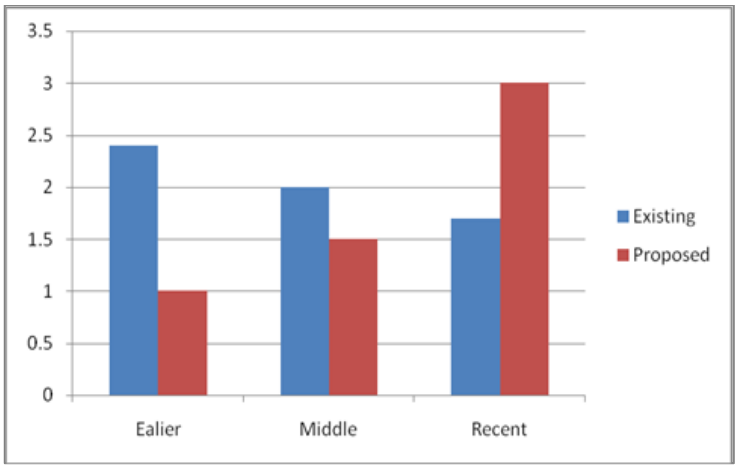

Figure 5. Comparative Anlaysis of Existing and Porposed Systems

5. Conclusion

Social networking sites are used to communicate with users and sharing information between the users in online. Millions of people are using the services of social networking sites like Facebook, twitter etc., User security is the main aspect in present days from person to person interaction. In this paper we propose to introduce efficient user security mechanism in transmission of messages from one user to another user with secret message sharing. For that we used filtering wall techniques with short text classifier users interact with the system by means of a GUI to set up and manage their FRs/BLs. The online social networks extend the process in different situations for accessing services from individuals in their user walls in same situational environment. Many empirical studies have shown that average OSN users have difficulties in understanding also the simple privacy settings provided by today OSNs. To overcome this problem, a promising trend is to exploit data mining techniques to infer the best privacy preferences to suggest to OSN users, on the basis of the available social network data. So we are proposed, decision tree algorithm to get efficient processing social networking sites. Moreover we can provide better security to social networking users.

References

[1] Denning, Dorothy E., "An intrusion-detection model", IEEE Transactions on Software Engineering, Volume. 2, 1987, pp. 222-232.
[2] K. Nirmala1, S. Satheesh kumar, Dr. J. Vellingiri, "A Survey on Text categorization in Online Social Networks**",** International Journal of Emerging Technology and Advanced Engineering, Volume. 3, Issue. 9, September 2013, pp. 446-450.
[3] N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: Two sides of the same coin?", Communications of the ACM, vol. 35, no. 12, 1992, pp. 29–38.
[4] P. J. Denning, "Electronic junk", Communications of the ACM, vol. 25, no. 3, 1982, pp. 163–165.
[5] P. W. Foltz and S. T. Dumais, "Personalized information delivery: An analysis of information filtering methods", Communications of the ACM, vol. 35, no. 12, 1992, pp. 51–60.
[6] P. S. Jacobs and L. F. Rau, "Scisor: Extracting information from online news", Communications of the ACM, vol. 33, no. 11, 1990, pp.87-97.
[7] S. Pollock, "A rule-based message filtering system", ACM Transactions on Office Information Systems, vol. 6, no. 3, july 1998, pp. 232–254.
[8] P. E. Baclace, "Competitive agents for information filtering", Communications of the ACM, vol. 35, no. 12, 1992, pp. 50.
[9] Zelikovitz, Sarah, and Haym Hirsh, "Improving short text classification using unlabeled background knowledge to assess document similarity", Proceedings of the seventeenth international conference on machine learning, June 29-July 02, 2000, pp. 1183-1190.
[10] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls", IEEE Transactions on Knowledge And Data Engineering, Vol. 25, No. 2, 2013, pp. 285-297.
[11] K. Strater and H.Richter, "Examining privacy and disclosure in a social networking community", Proceedings of the 3rd symposium on usable privacy and security, ACM, New York, USA, 2007,pp. 157-158.
[12] L. Fang and K. LeFevre, "Privacy wizards for social networking sites", Proceedings of the 19th international conference on World Wide Web (WWW 2010), New York, NY, USA: ACM, 2010, pp. 351–360.

Authors

Venkata Naresh Mandhala received M.Tech degree in Computer Science and Engineering from JNTU Hyderabad in 2012. He is a research student of VFSTR University. His research interests includes Image Processing, Data Mining, and Cloud Computing.

Debnath Bhattacharyya, Ph.D. (Tech, Computer Science and Engineering) from University of Calcutta, and M.Tech. (Computer Science and Engineering) from West Bengal University of Technology, Kolkata. Currently, he is associated as a Professor with Computer Application Department at RCCIIT, Kolkata. He has 18 years of experience in Teaching, and Research. His research interests include Bio-Informatics, Image Processing and Pattern Recognition. He has published 145 Research Papers in International Journals and Conferences and 4 Text Books for Computer Science.

Prof. Tai-hoon Kim, M.S., Ph. D (Electricity, Electronics and Computer Engineering), currently, Professor of Sungshin Women's University, Seoul, Korea. His research interests include Multimedia security, security for IT Products, systems, development processes, operational environments, etc. He has 20 Years of experience in Teaching & Research. He has already got distinctive Academic Records in international levels. He has published more than 250 Research papers in International & National Journals and Conferences.

