Volume 5, Issue 2, February 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Filtering System to Prevent Unwanted Messages in User Wall and Friend Recommendation in Online Social Network

T. Vijaya Lakshmi
PG Student, Department of CSE,
Sri Shakthi Institute of Engineering and Technology,
Coimbatore, Tamil Nadu, India

C. P. Shabariram
Assistant Professor, Department of CSE,
Sri Shakthi Institute of Engineering and Technology,
Coimbatore, Tamil Nadu, India

Abstract— Online social networks (OSNs) like, Google, Facebook and Twitter are designed to share public and personal information and make social relation with people who can be even strangers. One basic problem in today's OSN is to give users the ability to control messages shared on their own profile to avoid unwanted content to be displayed in their space. Till now, OSNs does'nt satisfies this requirement. This is gained by an automated system, which allows users to have direct control on their profile through the support of content-based filtering. The automated system also encourages friend recommendation, which is done based on the interest and lifestyle of the user.

Keywords— online social network, content-based filtering, friend recommendation.

I. INTRODUCTION

70% of the issues that arise in online social network is regarding the security of the data and privacy for user space. For the protection of user data, current OSNs indirectly require users to be system and policy administrators for regulating their data, where users can restrict data sharing to a specific set of trusted users. OSNs often use user relationship and group membership to distinguish between trusted and untrusted users. For example, in Facebook, users can allow friends, friends of friends (FOF), groups, or public to access their data, depending on their personal authorization and privacy requirements. Inspite of all these facilities that are provided by social networking sites, there are several problems that may arise. Any other user either trusted or untrusted can post unwanted message in particular user profile. This may affect the user's identity, name and fame in the social network. Next important issue in social network is in friend recommendation. Till now friend recommendation in social network is done only based on the locality of the user. For example, the institution the user studies or the organization the user work for or the place the user resides, etc. The major problem in this type of friend recommendation is the user cannot make friendship with any other user who does'nt match these criteria.

II. SYSTEM ANALYSIS

A. *Existing system*

Indeed, today OSNs provide very little support to prevent unwanted messages on user walls. For example, Facebook[7] allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. Friends are recommended to the user only based on few criteria like common friends, the organization they work for, the institution they studied, the place they belong to, etc., This is not an efficient friend recommendation system as the user is unaware of the person with similar taste.

B. *Proposed system*

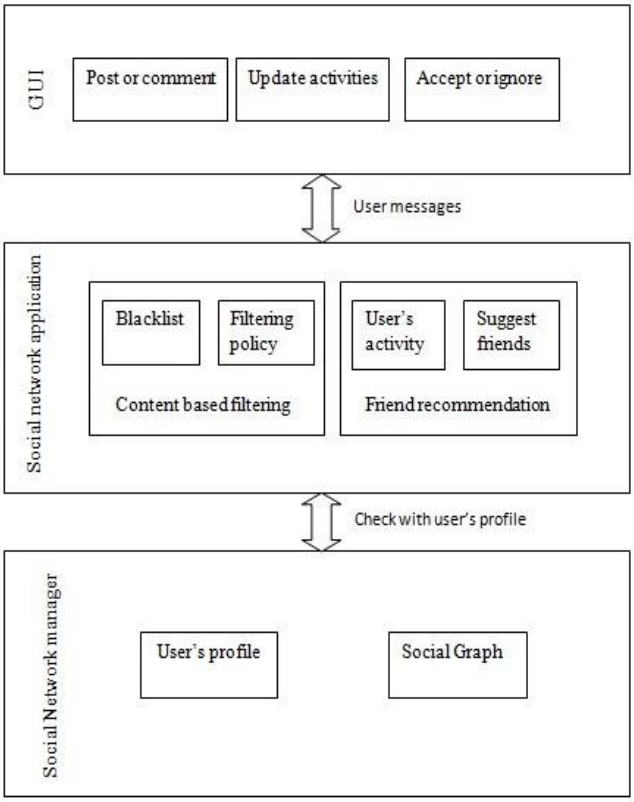
An automated system, called Filtered Wall (FW) is built, which is able to filter unwanted messages from OSN user walls. We use Machine Learning (ML) categorization of text method to assign automatically with each text message, a set of categories based on its content contained by it. The main effort in constructing a robust short text classifier (STC) is concentrated in the extraction and selection of a set of characterizing and discriminate features. Apart from classification facilities, the system gives a powerful rule layer exploiting a feasible language to specify Filtering Rules (FRs), by which users can specify what contents, cannot be displayed on their profile. FRs can support a variety of different filtering criteria that can be combined and customized according to the user needs. FRs exploit user profiles, user relationships as well as the output of the ML categorization process to state the filtering criteria to be enforced. In addition, the system provides the support for user-defined Blacklists (BLs), that is, lists of users that are temporarily prevented to post any kind of messages on a user wall.

For friend recommendation, the daily activities of the user are gathered from the status the user posts in the social network, for example watching movie, playing chess, etc., The recommendation system is built on the assumption that the user most probably posts messages on his/her wall with their own interest. For example, if a person is interested in

photography he/she might post more only regarding photography. The content of these posts are scanned and topic modeling[8] is done to choose the specific areas of interests. Then these areas are ranked. Ranking is done based on the frequency of the areas that occur in the content the user posts. The top ranked areas are selected and friends are recommended to the user in those areas.

C. System Architecture

The overall architecture of the automated system that is built is shown below. The architecture in support of OSN services is a three-tier structure. The first layer, called Social Network Manager (SNM), commonly aims to provide the basic OSN functionalities (i.e., profile and relationship management), whereas the second layer provides the support for external Social Network Applications (SNAs).4 The supported SNAs may in turn require an additional layer for their needed Graphical User Interfaces (GUIs)



III. RELATED WORK

A. Content Filtering

Content filtering[1], in the most general sense, involves using a program to prevent access to certain items, which may be harmful if opened or accessed. The most common items to filter are executables, emails or websites. Content filters can be implemented either as software or via a hardware-based solution. Content filtering works by matching strings of characters. When the strings match, the content is not allowed through. Content filters are often part of Internet firewalls. In such a usage content filtering is serving a security purpose, but content filtering is also used to implement company policies related to information system usage. For example, it's common to filter websites containing pornographic materials or social-networking sites unrelated to work.

Content-control software, content filtering software, secure web gateways, censor ware, Content Security and Control, web filtering software, content-censoring software, and content-blocking software[2] are terms describing software designed to restrict or control the content a reader is authorized to access, especially when utilized to restrict material delivered over the Internet via the Web, e-mail, or other means. The issue in these tools is that it cannot support in online social networks. It can only prevent the user from viewing unwanted content in their own computer.

B. Machine learning Classifier

Machine learning is a scientific practice that explores the building and study of algorithms that can be learnt from data.[3]Such algorithms operate by constructing a model based on inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions. Machine learning can be viewed as a subfield of computer science and statistics. It has strong relation with artificial intelligence and optimization, which deliver methods, theory and application domains to the field. Machine learning applied in a range of computing tasks where designing and programming explicit, rule-based algorithms is infeasible. Some applications include spam filtering, optical character recognition (OCR)[4] search engines and computer vision. Machine learning is sometimes conflated with data mining,[5] although that focuses more on exploratory data analysis. Machine learning and pattern recognition "can be viewed as two facets of the same field.

C. Friend Recommendation

Friend suggestion or friend recommendation is the general features that are available in online social network. Online social networks improve social experience by connecting users with common interests. Similar to real life, seeking good friends is much easier with recommendations in online social networks. Specially, a clear understanding is required, whether users who contribute more are more popular among other users, whether users like to make friends with popular users and the role difference of users with different diversity of individual interests in friendship formation[6]. A novel approach based on topic modeling is used to characterize the interest diversity degree of each user. The interest diversity features are used to help predict friend relationships between users.

D. Topic Modeling

In machine learning and natural language processing, a topic model[9] is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents. Intuitively, given that a document is about a particular topic, one would expect particular words to appear in the document more or less frequently: "dog" and "bone" will appear more often in documents about dogs, "cat" and "meow" will appear in documents about cats, and "the" and "is" will appear equally in both. A document typically concerns multiple topics in different proportions; thus, in a document that is 10% about cats and 90% about dogs, there would probably be about 9 times more dog words than cat words. A topic model captures this intuition in a mathematical framework, which allows examining a set of documents and discovering, based on the statistics of the words in each, what the topics might be and what each document's balance of topics is.

IV. MODULES

A. User Interface Design

User interface is the computers design, with the main focus on the interaction and experience of the user. The aim of user interface is to make user's interaction as simple as possible. The end user can execute the project easily with the help of GUI.

B. User Profile Creation

A user profile (user profile, or simply profile when used in-context) is a collection of personal data associated to a specific user. A profile refers therefore to the explicit digital representation of a person's identity. A user profile can also be considered as the computer representation of a user model. A user profile is a visual display of personal data associated with a specific user, or a customized desktop environment. A profile refers therefore to the explicit digital representation of a person's identity. A user profile can also be considered as the computer representation of a user model. A profile can be used to store the description of the characteristics of person. This information can be exploited by systems taking into account the persons' characteristics and preferences. The user personal data store in ONLINE social networks (OSNs) database that details contain informs like first name, last name, username, password, email Id, gender etc.

C. Post Wall Creation

The Website wallpost in most social network is enabled with photo sharing activities. Protected albums allow users to set their albums with access protection. This is one of the beneficial features from wallpost that who fear with photo scams on photo sharing websites. Photo tagging the option makes the photo search easier after a long period of time. Although OSNs currently provide simple access control mechanisms allowing users to govern access to information contained in their own spaces, users, unfortunately, have no control over data residing outside their spaces.

D. Content Filtering

The training stage of the spam detector includes following steps:

Step 1: Training Set is divided into positive set (spam comments) and negative set (ordinary comments).

Step 2: The stop words for example prepositions, articles, etc., are removed.

Step 3: Stemming algorithm is applied which will produce all the available words after the removal of stop words in the form of noun.

Step 4: The remaining words after applying stemming algorithm, are checked for the presence of any spam messages.

Step 5: If the presence of spam is detected, the message will not get posted on the user wall.

E. Friend Recommendation

The content of the user profile are scanned and then the contents are classified according to the areas. Then the areas are ranked based on the probability of their frequency. The top most ranks are considered and the friend suggestion is done based on the similarity of the ranks.

V. CONCLUSION

An automated system called Filtered Wall is being built. This system is designed to automatically prevent the unwanted messages which are posted in one's profile. This is done by specifying the default filtering criterias. Additionally the friend recommendation is also performed where the friends are suggested to the users based on the interests and lifestyle rather than the locality of the users.

ACKNOWLEDGMENT

It's a great pleasure to thank my project guide Mr.C.P.Shabariram, Assistant Professor, Department of Computer Science and Engineering for her valuable technical suggestion continuous guidance throughout this project work.

REFERENCES

[1] http://www.techopedia.com/definition/14314/contentfilterig
[2] http://en.wikipedia.org/wiki/Content-control_software
[3] Ron Kovahi; Foster Provost (1998). "Glossary of terms". Machine Learning **30**: 271–274.
[4] Wernick, Yang, Brankov, Yourganov and Strother, Machine Learning in Medical Imaging,IEEE Signal Processing Magazine, vol. 27, no. 4, July 2010, pp. 25-38
[5] Mannila, Heikki (1996). "Data mining: machine learning, statistics, and databases". Int'l Conf. Scientific and StatisticalDatabase Management. IEEE Computer Society.
[6] When Diversity Meets Speciality: Friend Recommendation in Online Social Networks, Hao Wu†, Vikram Sorathia, Viktor K. Prasanna †Department of Computer Science Ming Hsieh Dept. of Electrical Engineering University of Southern California
[7] WWW.FACEBOOK.COM
[8] Zhibo Wang, Student Member, IEEE, Jilong Liao, Qing Cao, Member, IEEE, Hairong Qi,Senior Member, IEEE, and Zhi Wang, Member, IEEE, "Friendbook: A Semantic-based Friend Recommendation System for Social Networks"
[9] Blei, David M. (April 2012). "Introduction to Probabilistic Topic Models" (PDF). *Comm. ACM* **55** (4):84. doi:10.1145/2133806.2133826

