

CONTENT BASED FILTERING WITH BLACKLIST

By

B.Vinitha, S.Sowmiya, R.Uma

Lecturer, Department of IT, PSG Polytechnic College, Coimbatore, India
uma.dit@psgpolytech.ac.in

ABSTRACT

One fundamental issue in today's online social networks (OSNs) is to give users the ability to control the messages posted on their own private space and to avoid the unwanted content that is displayed. Right now, OSNs provide only little support to this requirement. To fill the gap, a system is proposed which allows OSN users to have a direct control on the messages posted on their private space. This is achieved through a flexible rule-based system that allows the users to customize the filtering criteria to be applied to their space, and a Machine Learning-based soft classifier that automatically labels messages in support of content-based filtering. Short-text classifier is used which includes two levels; in the first level short-texts are labeled with crisp Neutral & non-neutral labels. The second level soft classifier acts on the non-neutral short texts, produces estimated appropriateness or "Gradual Membership" for each of the conceived classes, this list of grades is then used by the subsequent phases of the filtering process.

Keywords - Information Filtering, Short-text Classification, Blacklists and Policy-based Personalization.

1. INTRODUCTION

Online Social Networks (OSNs) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. Daily and continuous communications imply the exchange of several types of content, including free text, image, audio, and video data. In the sharing in OSNs, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact that in OSNs there is a possibility of posting or commenting other posts on particular public/private areas, called in general walls. Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages.

2. EXISTING SYSTEM

Face book allows users to specify who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends).

However, there is no restriction on the content of the message posted and therefore it is not possible to prevent the unwanted messages, such as political or vulgar ones, no matter of the user who posts them.

A. Drawback of Existing System

Even though, the social networks have restrictions on the users who can post messages and comment on any other user's wall they don't provide any restriction on what they post.

Since wall messages are composed of short text the traditional classification methods have limitations. It requires designing of ad hoc classification strategies.

3. ALGORITHM

A. Machine Learning (ML) Text Categorization Techniques

1. Short text classifier

i) Text Representation

We consider three types of features, Bag of Words (BOW), Document properties (DP) and Contextual Features (CF). They are entirely derived from the information contained within the text of the message. Text representation using endogenous knowledge has a good general applicability; however, in operational settings, it is legitimate to use also exogenous knowledge, i.e., any source of information outside the message body but directly or indirectly related to the message itself.

ii) Machine Learning-Based Classification

We address short text categorization as a hierarchical two level classification process. The first-level classifier performs a binary hard categorization that labels messages as Neutral and Non-neutral. The first level filtering task facilitates the subsequent second level task in which a finer-grained classification is performed. The second-level classifier performs a soft-partition of Non-neutral messages assigning a given message a gradual membership to each of the non-neutral classes.

iii) Radial Basis Function Networks (RBFN)

RBFNs have a single hidden layer of processing units with local, restricted activation domain: a Gaussian function is commonly used, but any other locally tunable function can be used.

B. Filtering rules and blacklist management

i) Filtering Rules:

Filtering rules (FRs) are the set of rules, specifies what content should be displayed by the user. In FR we consider some issues that affect a message filtering decision. While communicating the same message can have different meaning which depends on the user who writes it.

In FRs there should be constraints on message creators. Based on different criteria the creator should be selected. One of them is using conditions on user's profile attributes.

ii) Blacklists

A blacklist rule is a tuple (author, creatorSpec, creatorBehavior, T),

where

- author is the OSN user who specifies the rule, i.e., the wall owner;
- creatorSpec is a creator specification
- CreatorBehavior consists of two components RFBlocked and minBanned.
- T denotes the time period the users identified by creatorSpec and creatorBehavior.

4. SYSTEM ARCHITECTURE

The OSN has three layers; there are graphical user interface, social network application and social network managers. The social network managers handle the basic functionalities like Profile management, network based function etc. Application layer has short text classifier and content based message filtering. Short text classifier classifying the messages based on the content. Content based message filter has black list and filtering policies. First, find relationship between the user and it will filter and calculate the probabilities using classifier. And then send an empty message below the probabilities result to the user. So the proposed system will give the direct control to the user that what kind of messages are displayed on their wall.

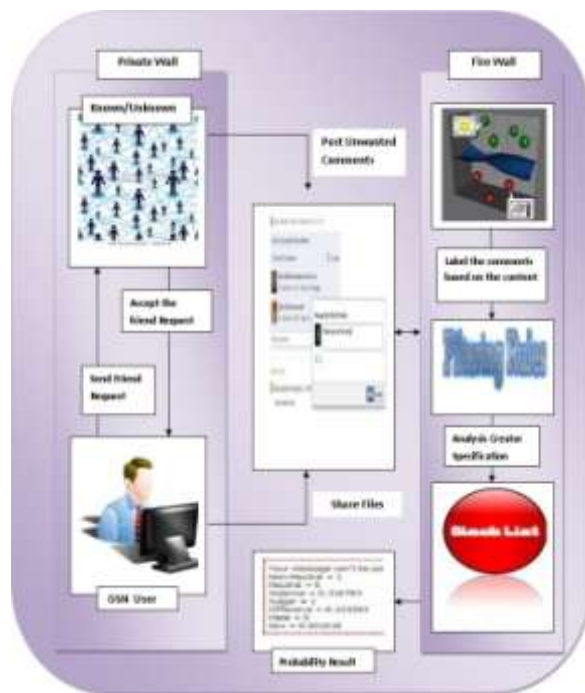


Fig 1: OSN Architecture

Proposed model

We propose and experimentally evaluate an automated system, called Filtered Wall, to filter the unwanted messages from Online Social Network user walls.

We exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content. The system provides a powerful rule layer exploiting a flexible language to specify Filtering Rules (FRs), by which users can state what contents, should not be displayed on their walls.

Authentication

A new user going to consume the service has to register first by providing necessary details. After successful completion of the sign up process, the user has to login into the application by providing username and exact password which was provided at

the time of registration.

Profile Generation

User's profile details are stored in the database which includes name, contact number, email id, photos, and other information. Logged users can see their details and if they wish, they can edit their profile.

Accept Friend Request

User can add new friends and view their details. Logged users can also view their friends list.

Send Request

Users can select friend to send a request. Logged user can accept the request.

Post Messages

User can post any photo in public wall, and anyone can post a comment for that photo. Unknown persons also post their comments who are not the members of the group.

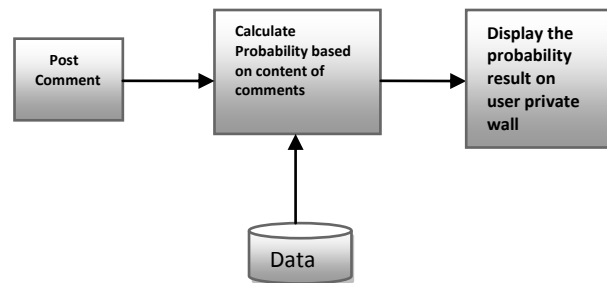


Fig 2: Posting comments

Block Unwanted Message

The probability of the message contents has to be calculated. That result will be displayed on the user's private wall with two options like accept and reject. If reject is chosen it will take to main page else it will remain in the login page itself.

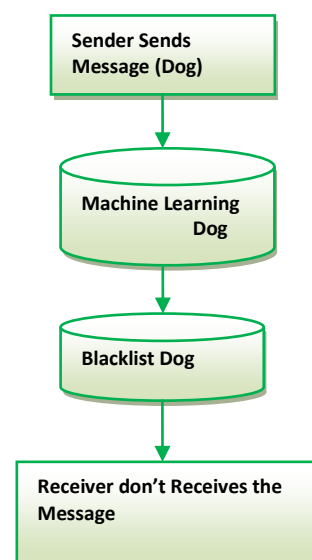


Fig 3: Blocking unwanted message

5. EXPERIMENTAL EVALUATION



Fig 4: Home Page



Fig 5: Registration Page



Fig 6: View the friend Request

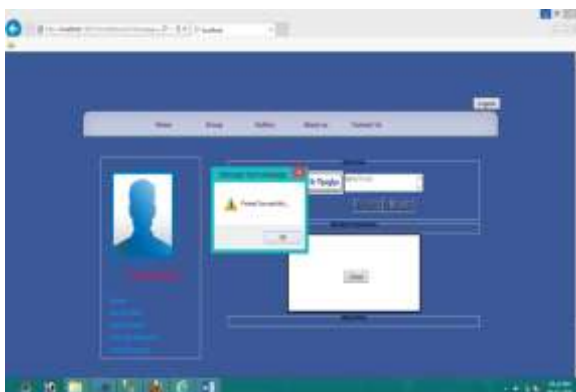


Fig 7: Post message



Fig 8: List the message details



Fig 9: Sender is blocked

6. CONCLUSION

A system that has direct control to block unwanted messages on the user's social network wall is developed using the machine language soft classifier and then applying the Filter Rule based on the creators. Whenever an unwanted message is posted that particular user who posts that message will be blocked. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of BLs. In future, message filtering can also be addressed by investigating the use of online learning paradigms which are able to include label feedbacks from the users.

7. REFERENCES

- [1] M. Chua and H. Chen, Decision Support Systems, vol. 44, no. 2, pp. 482-494, 2008 A Machine Learning Approach to Web Page Filtering Using Content and Structure Analysis.
- [2] R.J.Mooney and L.Roy, Proc.Fifth ACM Conf.Digital Libraries,pp.195-204,2000. Content-Based Book Recommending Using Learning for Text Categorization.
- [3] F. Sebastian, ACM Computing Surveys, vol. 34, no.1, pp.1-47, 2002 Machine Learning in Automated Text Categorization.
- [4] M. Vanzetti, E. Bengali, B. Carminative, M. Carlo, and E. Ferrari, Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010. Content-Based Filtering in On-Line Social Networks.
- [5] N.J. Belkin and W.B. Croft, Comm. ACM, vol. 35, no. 12, pp. 29-38, 1992. Information Filtering and Information Retrieval: Two Sides of the Same Coin?
- [6] P.J. Denning, Comm. ACM, vol. 25, no. 3, pp. 163-165, 1982 Electronic Junk.

[7] P.W. Foltz and S.T. Dumas, Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992 Personalized Information Delivery: An Analysis of Information Filtering Methods.

[8] P.S. Jacobs and L.F. Rau, Comm. ACM, vol. 33, no. 11, pp. 88-97, 1990 Scisor: Extracting Information from On-Line News.

[9] S. Pollock, ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988 A Rule-Based Message Filtering System.

[10] P.E. Baclace, Comm. ACM, vol. 35, no. 12, p. 50, 1992 Competitive Agents for Information Filtering.