Filter Wall : To prevent undesired messages posted on OSN user wall

Amruta Kachole,S.D.Jondhale

ME(computer), amrutakachole@gmail.com

Abstract— This paper is support for content based user preferences. It is possible to the use of a Machine Learning (ML) text categorization procedure able to automatically assign with each message a set of categories based on its content. The proposed approach is a key service for social networks where users have little control on the messages displayed on their walls. For Instance, Facebook 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. For instance, it is not possible to prevent political or vulgar messages. In contrast, by means of the proposed mechanism, a user can specify what contents should not be displayed on his/her wall, by specifying a set of filtering rules. Filtering rules are very flexible in terms of the filtering requirements they can support, in that they allow to specify filtering conditions based on user profiles, user relationships as well as the output of the ML categorization process. In addition, the system provides the support for user-defined blacklist management, that is, list of users that are temporarily prevented to post messages on a user wall.

Keywords— Online social networks, content based filtering, short text classification, Space Vector Model (SVM)

INTRODUCTION

On-line Social Networks (OSNs) have become a popular interactive medium to communicate, share and disseminate a considerable amount of human life information. Daily and continuous communication implies the exchange of several types of content, including free text, image, audio and video data. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data and then provide an active support in complex and sophisticated tasks involved in social networking analysis and management. The main part of social network content is constituted by short text, a notable example are the messages permanently written by OSN users on particular public/private areas, called in general walls.

In this paper an automated filtering system is implemented for Content based filtering that allows OSN users to have a direct control on the messages posted on their walls. This proposed approach can automatically filter unwanted messages from OSN user walls on the basis of content of message. It also proposes a flexible rule-based system that allows users to customize the filtering criteria to be applied to their walls and a Machine Learning-based soft classifier automatically labeling messages in support of content-based filtering. The core components of the proposed system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier modules. The short text classifier component aims to classify messages according to a set of categories. STC is performed 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 second-level classifier performs a soft-partition of Non-neutral messages assigning a gradual membership to each of the non-neutral classes. Therefore, ML-based short text classifier extracts metadata from the content of the message In contrast, the Content-Based Messages Filtering component exploits the message categorization provided by the STC module to enforce the FRs specified by the user. BLs can also be used to enhance the filtering process.

Modules

Module 1: Vector Presentation

Module 2: Binary Classification

Module 3: Multi Label Classification

Module 4: Filtering Rules Specification

Module 5: Blocking Management

Module 1: Vector Representation

In this project, training set are prepared from the data set, WmSnSec which is available online at http://www.dicom.uninsubria.it/~marco.vanetti/wmsnsec. Training set is divided into neutral and non neutral training dataset to train the SVM classifier. SVM is trained by extracting the content of messages in the dataset (wwsnsec). This approach follows Vector Space model, according to which a text message dj is represented as a vector of binary or real weights $d_j = \{w_{1j,} w_{2j,\ w3j\ \ldots\ldots}\ w_{|T|j\}}$ where T is the set of terms that occur at least once in at atleast one document of the collection T_r and $w_{kj} \in [0;1]$ represents that how much the term k contributes to the semantics of the document. Training (T_rS_D) set are transformed into a form of vector representation

$$T_rS_D = \{ (\vec{x}_i, \vec{y}_i) \ldots\ldots\ldots\ldots (\vec{x}_{|TrSD|}, \vec{y}_{|TrSD|}) \}$$

Test set (T_eS_D) are transformed into a vector representation as &.

$$T_eS_D = \{(\vec{x}_i, \vec{yi}) \ldots\ldots\ldots\ldots (\vec{x}_{|TeSD|}, \vec{y}_{|TeSD|}) \}$$

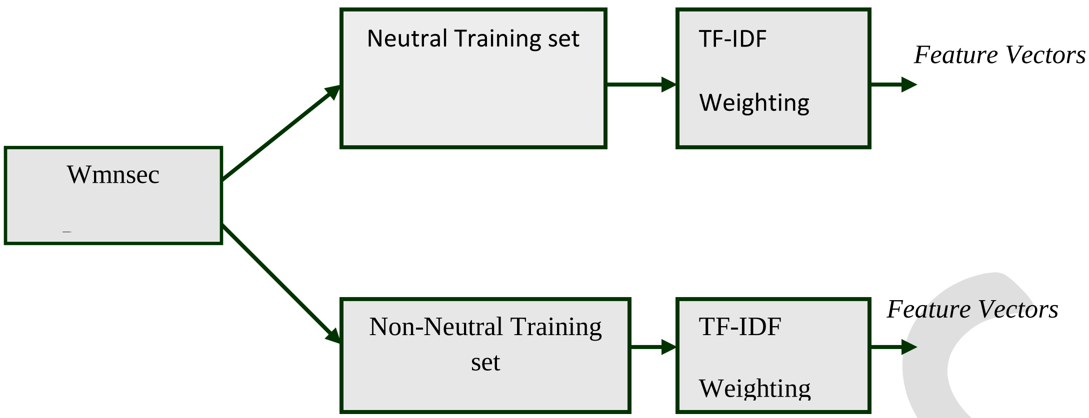
The term frequency-inverse document frequency is used to calculate the weight of term tk in document dj as follows

Tf-idf weighting = # (t_k, dj) * log N/ # (tk, N)

(t_k, dj) is the term frequency where the number of occurrences of term tk in the document dj.

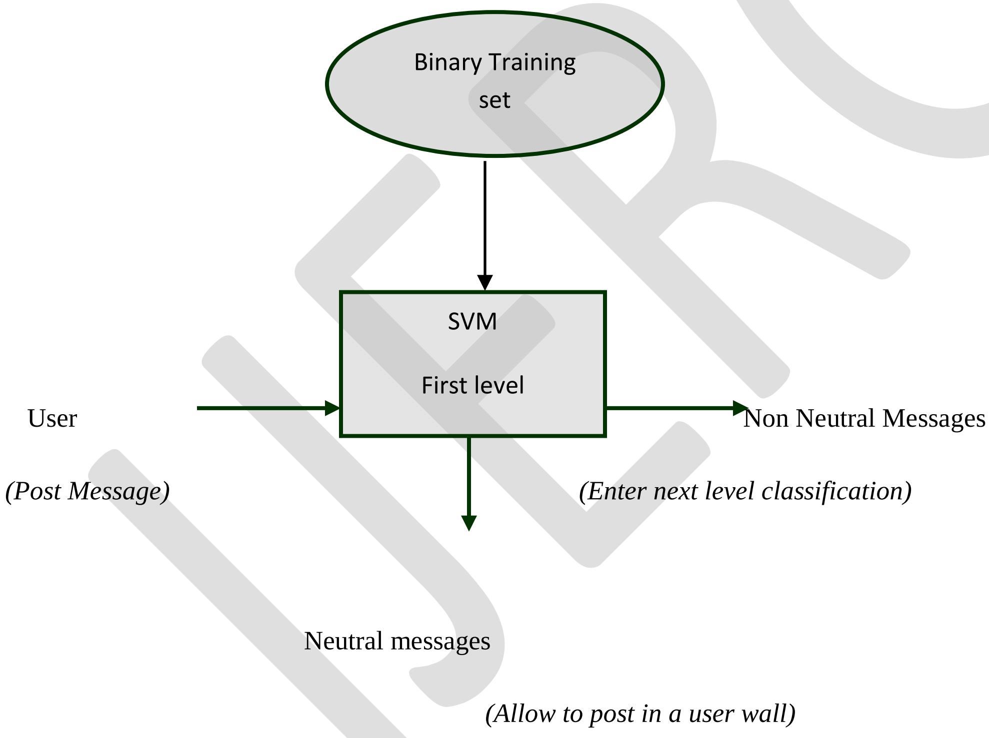
log N/ # (tk, N) is the inverse document frequency i.e., document frequency the number of documents have the term tk among all the documents N

ISSN 2091-2730



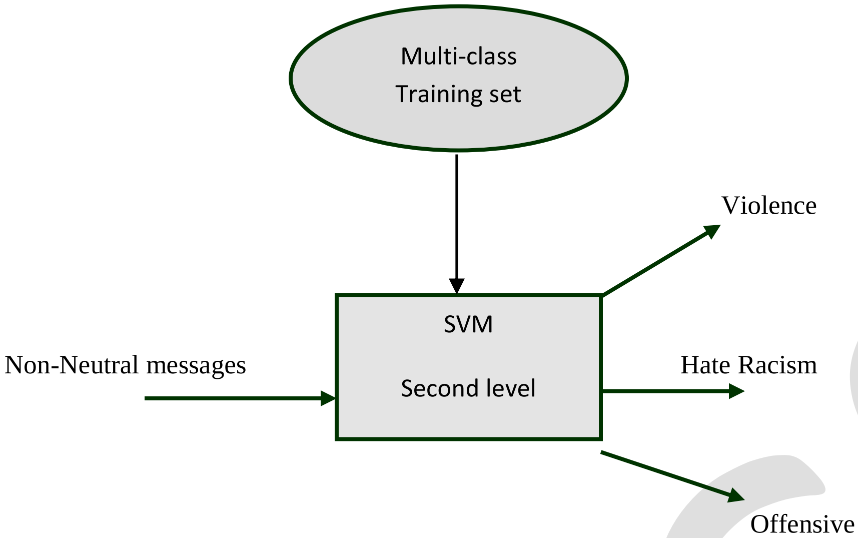
Module 2: First level Binary classification

In this project, SVM perform two level of classification to filter the unwanted short text based on its content. Let m1 be the first level classifier used to classify the messages into two types such as Neutral and Non neutral messages.



Module 3: Multi Label Classification

In this module, the messages which are labeled as non neutral messages are given as an input to the second classifier M2 that performs multi-label classification. In order to perform classification, the classifier M_2 is trained using multi-class training set as follows . The performance of the model M2 is then evaluated using the test set TeS2.



Module 4: Filtering Rule Specification

Besides classification facilities, this project provides a powerful rule layer exploiting a flexible language to specify Filtering Rules (FRs), by which users can state what content should not be displayed on their walls. FRs supports the specification of content-based filtering using variety of different filtering criteria in order to combine and customize the user needs. More precisely, FRs exploit user profiles, user relationships as well as the output of the ML categorization process to state the filtering criteria to be enforced. FRs should allow users to state constraints on message creators. This implies to state conditions on type, depth, and trust values of the relationship(s) creators should be involved in order to apply them the specified rules. A Filtering Rule FR is represented as a tuple as follows

FR = (author, creatorSpec, contentSpec, action)

→ author is the user who specifies the rule;

→ creatorSpec is a creator specification,

→ contentSpec is a Boolean expression that expresses constraints in the form (c, ml) where C is a class of the first or second level and ml is the minimum membership level threshold required for class C to make the constraint satisfied.

→ action ∊ {block; notify} denotes the action to be performed by the system on the messages matching contentSpec and created by users identified by creatorSpec.

Module 5: Block List Management

Block List management is used to avoid messages from undesired creators, independent from their contents. To achieve this, user specifies the information through a set of rules called as BL rules. These rules are directly managed by the system, which determine who are the users will be inserted into the Block List and decide when user retention in the BL is finished. A BL rule is a tuple {author, creatorSpec, creatorBehavior, T} , where creatorBehavior consists of two components RFBlocked and minBanned. RFBlocked (RF, mode, window) is defined such that

$$\text{Relative Frequency} = \frac{\#bMessages}{\#tMessages},$$

→ #tMessages is the total number of messages that each OSN user identified by creatorSpec has tried to publish in the author wall (mode = myWall) or in all the OSN walls (mode = SN); whereas #bMessages is the number of messages among those in #tMessages that have been blocked;

→ window is the time interval of creation of those messages that have to be considered for RF computation;

→ minBanned ¼ (min, mode, window), where min is the minimum number of times in the time interval specified in window that OSN users identified by creatorSpec have to be inserted into the BL due to BL rules specified by author wall (mode = myWall) or all OSN users (mode = SN) in order to satisfy the constraint.

→ T denotes the time period the users identified by creatorSpec and creatorBehavior have to be banned from author wall.

CONCLUSION

In OSN environment, the privacy preservation for data analysis, share and mining is a challenging research issue due to the difficulties in traditional classification approaches, thereby requiring intensive investigation. This project presented a system to filter undesired messages from OSN walls. This project improved the quality of classification using short text classifier. It provides high privacy and flexibility to manage OSN walls. In this project, we have investigated the privacy problem of user by flexible filtering rule specification and designed a group of hierarchical level classification to assign the metadata for each of the message posted by the user. This project creatively specified flexible rule specification to directly control the messages posted on their walls. It concretely accomplishes the automated filtering in a highly flexible way.

REFERENCES:

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo ”A System to Filter Unwanted Messagesfrom OSN User Walls”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 2, FEBRUARY 2013 285

[2] A. Adomavicius and G. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,” IEEE Trans. Knowledge and Data Eng., vol. 17,no. 6, pp. 734-749, June 2005.

[3] M. Chau and H. Chen, “A Machine Learning Approach to Web Page Filtering Using Content and Structure Analysis,” Decision Support Systems, vol. 44, no. 2, pp. 482-494, 2008.

[4] R.J. Mooney and L. Roy, “Content-Based Book Recommending Using Learning for Text Categorization,” Proc. Fifth ACM Conf.Digital Libraries, pp. 195-204, 2000.

[5] F. Sebastiani, “Machine Learning in Automated Text Categorization,” ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[6] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari,“Content-Based Filtering in On-Line Social Networks,” Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML ’10), 2010.

[7]] N.J. Belkin and W.B. Croft, “Information Filtering and Information Retrieval: Two Sides of the Same Coin?” Comm. ACM, vol. 35,no. 12, pp. 29-38, 1992.

[8]] P.J. Denning, “Electronic Junk,” Comm. ACM, vol. 25, no. 3,pp. 163-165, 1982.

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

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

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

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

[13] P.J. Hayes, P.M. Andersen, I.B. Nirenburg, and L.M. Schmandt,"Tcs: A Shell for Content-Based Text Categorization," Proc. Sixth IEEE Conf. Artificial Intelligence Applications (CAIA '90), pp. 320-326, 1990.

[14]] G. Amati and F. Crestani, "Probabilistic Learning for Selective Dissemination of Information," Information Processing and Management,vol. 35, no. 5, pp. 633-654, 1999.