ISSN: 2320-1363

Neural Network classifier to filter unwanted messages from OSN user walls

P.Narendra *1, T.Subhashini *2, R.Pitchaiah *3

1*M.Tech Scholar, Dept of CSE, Universal College of Engineering & Technology, Perecherla, Dist: Guntur, AP, India

2*Assistant Professor, Dept of CSE, Universal College of Engineering & Technology, Perecherla, Dist: Guntur, AP, India

3*Associate Professor & HOD, Dept of CSE, Universal College of Engineering & Technology, Perecherla, Dist: Guntur, AP, India

Abstract:

Online social networks play an important role in personnel and business life of many individuals. The rapid usage of OSN leads to many issues like privacy, misleading and unrelated information/messages on user's walls. In this paper we are addressing an issue of unwanted messages. We are using rule based system to customize the user's messages on the walls and a neural network classifier to filter and remove unwanted messages automatically using content-based filtering.

Keywords: online social networks, message filtering, neural network classifier, rule-based system.

1. Introduction

Online Social Networks plays an important in the usage of internet applications. Millions of people are using Online Social networks like Facebook, Twitter, Linked in etc. These applications improve the social and public relationships. Each user can maintain his personnel information like date birth, interests, hobbies, current location, job and so on. This information is completely private to the user and some part of information can be visible to his friends, and relations. User can share some data like status, images and any content on his friends and public profile. By using this public information and information of users friends of friends the attackers can be get the personnel information of the user. The gathered information can be misused by attackers to publish adds etc.

Along with privacy controlling information on user's walls is another major issue in Online Social Networks. There is a

ISSN: 2320-1363

very high chance of posting unwanted content on particular public/private areas, called in general walls. So, to control this type of activity and prevent the unwanted messages which are written on user's wall we can implement filtering rules (FR) in our system. Also, Black List (BL) will maintain in this system. OSNs provide support to prevent unwanted messages on user walls. For example, 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 and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them.

2. Related work:

We are proposing a new online social network to customize and filter the unwanted messages using content based filtering. There is a lot research going on this issue. The work mainly classified into three ways.

2.1 Content based filtering
2.2 Collaborative filtering
2.3 Policy based filtering

2.1 Content based filtering

In content-based filtering, each user is assumed to operate independently. As a result, a content-based filtering system selects information items based on the correlation between the content of the items and the user preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences [1], [2].

Content-based filtering is mainly based on the use of the ML paradigm according to which a classifier is automatically induced by learning from a set of pre classified examples. A remarkable variety of related work has recently appeared, which differ for the adopted feature extraction methods, model learning, and collection of samples [3], [4],[5], [6], The feature extraction procedure maps text into a compact representation of its content and is uniformly applied to training and generalization phases. Several experiments prove that Bag-of-Words (BoW) approaches yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statistical quality [7], [8]. As far as the learning model is concerned, there are a number of major approaches in content-based filtering and text classification in general showing mutual advantages and disadvantages in function of application dependent issues. In [10], a detailed comparison analysis has been conducted confirming superiority of Boosting-based classifiers [11], Neural Networks [12], [13], and Support Vector Machines [14] over other popular methods, such as Rocchio [15] and Naıve Bayesian . However, it is worth to note that most of the work related to text filtering by ML has been applied for long-form text and the assessed performance of the text classification methods strictly depends on the nature of textual documents.

2.2 Collaborative filtering

In collaborative filtering information will be selected on the basis of user's preferences, actions, predicts, likes, and dislikes. Match all this information with other users to find out similar items. Large dataset is required for collaborative filtering system. According to user's likes and dislikes items are rated.

2.3 Policy based filtering:

ISSN: 2320-1363

In policy based filtering system users filtering ability is represented to filter wall messages according to filtering criteria of the user. Twitter is the best example for policy based filtering [16] In that communication policy can be defines between two communicating parties.

3. Proposed System:

We are proposing new way of blocking the unwanted messages on users walls by using neural network classifier and filtering rules. The proposed system contains the following modules called

3.1 Short text classifier

3.2 Blocking the unwanted messages

3.1 Short text classifier

Aim of the short text classifier is to recognize and eradicate the neutral sentences and categorize the non neutral sentences in step by step, not in single step. This classifier will be used in hierarchical strategy. The first level task will be classified with neutral and non neutral labels. The second level act as a non neutral, it will develop gradual membership. These grades will be used as succeeding phases for filtering process. Short text classifier includes text representation and neural network classification.

Text representation:

Representing the text of a document is critical,which will affect the classification performance. Many features are there for representation of text ,but we judge three types of features. BOW, Document properties (DP) and contextual features. BOW and Document properties are already used in[17], are endogenous that is , text which is entirely derived from the information within the text message. Endogenous knowledge is well applicable in representation of text. It is genuine to use also exogenous knowledge in operational settings. Exogenous knowledge is termed as any source of information from outside the message but directly or indirectly communicate to the message itself. CF modeling is introduced; its feature is to understand the semantics of message. DP features are heuristically evaluated. some domain specific criteria is considered, trial and error procedures are needed for some cases. Some of them are,

Correct words: It states the amount of terms. Correct words will be calculated.

Bad words: comparison to the correct words will be evaluated. collection of dirty words will be determined.

Capital words: It will say about the amount of words written in message. Percentage of words in capital case will be calculated.

Punctuations characters: Percentage of punctuation character over the total number of character will be calculated.

Exclamation mark: Percentage of exclamation marks over the total number of punctuation characters will be calculated.

Question marks: Percentage of question marks over the total number of punctuation character will be evaluated.

Neural Network Classifier

The neural network classifier has two stages. First stage involves the training of the classifier to detect unwanted messages. The second stage testing and identifying the unwanted messages. We are using supervised learning algorithm to detect error rate in back propagation algorithm to update the weights during traing. We are maintaining set of input and output words as training patterns like {(kill,bad word),(good, good word),(xxx,bad word)……..}. This is our training dataset updates when a new word comes. The final

ISSN: 2320-1363

set of updated weights is obtained after certain no of iterations. These weights are used to for testing and blocking the unwanted messages by social network manager.

3.2 Blocking the unwanted messages:

This is another important module of our proposed system which uses the neural network classifier to classify the unwanted or bad messages which should not be posted on the user's walls. Even the messages sent by friends also will be blocked if the message contains more bad words.

4. Algorithm:

Input: Messages, training data set

Output: Blocking messages

Algorithm:

Step1: remove unnecessary words like punctuation marks, exclamation mark, Question mark etc.

Step 2: Train the neural network classifier to classify the bad words and good words. And obtain the weights for testing any images.

Step3: Block the message by using the neural network classifier.

5. Conclusion:

Our proposed system dynamically blocks the unwanted messages on user's wall in Online Social networks. Even thought it looks simple technique the performance of blocking system improves. The efficiency of the classifier depends on the training data set. We have include the more words to get most optimal results. We can extend our work for filtering the unwanted images too.

References:

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

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

[3] 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.

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

[5] M.J. Pazzani and D. Billsus, "Learning and Revising User Profiles:The Identification of Interesting Web Sites," Machine Learning,vol. 27, no. 3, pp. 313-331, 1997.

[6] Y. Zhang and J. Callan, "Maximum Likelihood Estimation for Filtering Thresholds," Proc. 24th Ann. Int'l ACM SIGIR Conf.Research and Development in Information Retrieval, pp. 294-302, 2001.

[7] C. Apte, F. Damerau, S.M. Weiss, D. Sholom, and M. Weiss,"Automated Learning of Decision Rules for Text Categorization,"Trans. Information Systems, vol. 12, no. 3, pp. 233-251, 1994.

[8] S. Dumais, J. Platt, D. Heckerman, and M. Sahami, "Inductive Learning Algorithms and Representations for Text Categorization,"Proc. Seventh Int'l Conf. Information and Knowledge Management (CIKM '98), pp. 148-155, 1998.

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

[10] R.E. Schapire and Y. Singer, "Boostexter: A Boosting-Based System for

ISSN: 2320-1363

Text Categorization," Machine Learning, vol. 39,nos. 2/3, pp. 135-168, 2000.
[11] H. Schu¨ tze, D.A. Hull, and J.O. Pedersen, "A Comparison of Classifiers and Document Representations for the Routing Problem," Proc. 18th Ann. ACM/SIGIR Conf. Research and Development in Information Retrieval , pp. 229-237, 1995.
[12] E.D. Wiener, J.O. Pedersen, and A.S. Weigend, "A Neural Network Approach to Topic Spotting," Proc. Fourth Ann. Symp.Document Analysis and Information Retrieval (SDAIR '95), pp. 317-332, 1995.
[13] T. Joachims, "Text Categorization with Support Vector Machines:Learning with Many Relevant Features," Proc. European Conf.Machine Learning, pp. 137-142, 1998.
[14] T. Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," Proc. Int'l Conf. Machine Learning, pp. 143-151, 1997.
[15] S.E. Robertson and K.S. Jones, "Relevance Weighting of Search Terms," J. Am. Soc. for Information Science, vol. 27, no. 3, pp. 129-146, 1976.
[16] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo," A System to Filter Unwanted Messages from OSN User Walls" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 2,FEBRUARY 2013.
[17] B.Sriram, D.Fuhry, E.Demir, H.ferhatatosmanoglu, and M.Demirbas, "Short Text Classification in Twitter to Improve Information Filtering," Proc.33rd Int'l ACM SIGIT Conf. Research and Development in Information Retrieval(sIGIR '10), pp.841-842,2010.