International Journal of Advance Research in Science and Engineering
Vol. No. 5, Issue No. 08 , August 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

PERFORMANCE EVALUATION OF RADIAL BASIS FUNCTION NETWORK AND NAIVE BAYES FOR FILTERING OF UNWANTED MESSAGES FROM OSN USER WALLS

Rupal B. Jain[1], D. R. Patil[2]

[1,2]*Department of Computer Engineering, R.C.P.I.T., Shirpur, (India)*

ABSTRACT

On-line Social Networks (OSNs) are one of the most popular interactive medium to share and communicate among the internet user. In OSN user may post or commenting on others wall. Up to now, OSN give very little support to avoid unwanted messages displaying on user walls. There is no support for content-based preferences. So to fill this gap, in this paper we have used a flexible rule-based system in which we have applied different filtering rules to filter incoming messages on user's wall and classifier to automatically label the messages in support of content based filtering. In this paper, we have used Radial Basis Function Network and Naive Bayes classifier for text classification. We have compared the performance of this two classifier, experimental result shows that Naive Bayes take less learning time as compare to Radial Basis Function Network.

Keywords : On-line Social Networks, Information Filtering.

I. INTRODUCTION

On-line Social Networks (OSNs) gives a platform to communicate, share human life information. Now a day's popularity of OSN increases day by day. Daily and continuous communication may include different type of data like text, audio, video etc [1]. Social network services consist of each user having his/her own profile, his/her social link or different types of services. It is web-based service which allows user to create own profile, to connect with different user to communicate with each other. They may post different messages on other private/public wall. Therefore in OSN, there is chance of posting unwanted content on others wall. Up to now OSNs provide very little support to avoid unwanted message displaying on the users wall. Existing system does not support content-based preferences, so it is not possible to avoid unwanted messages, such as vulgar, political or etc. Our main aim is to provide ability to user to automatically control the messages written on their own wall by avoiding unwanted messages. But generally this wall messages are very short i.e. they do not give sufficient word occurrence and traditional classification method gives limitation for short text. So we analyze two methods for text classification i.e. Radial Basis Function Network and Naive Bayes classifier to automatically categorize each text messages based on its content.

With the classification facilities, the system also provides powerful Filtering rules (FRs) with flexible language by which user can state which content should not displayed on the user wall. According to the user need FRs can have different filtering criteria. In addition with FRs system also gives user-defined BlackLists (BLs) i.e. it gives the list of user which are temporarily banned to post messages within particular time period.

The aim of the present work is to give a automated system, called Filtered Wall (FW), which is able to filter unwanted messages from OSN user walls. And by using classifier automatically assign a set of categories based on its content to each text. Also we have compared two classifier Radial Basis Function Network and Naive Bayes, experimental result shows that Naive Bayes take less learning time as compare to Radial Basis Function Network.

II. RELATED WORK

M.Chau et al. proposed a machine-learning-based approach that combines Web structure analysis and Web content analysis. They represent each web page with a set of link-based and content-based features, which is used as the input for different types of machine learning algorithms. Their approach was implemented using both support vector machine and feedforward/backpropagation neural network. The neural network approach (NN-WEB) and support vector machine approach (SVM-WEB) was compared against two benchmark approaches:

1) a lexicon-based approach (LEXICON), and
2) a keyword-based support vector machine approach (SVMWORD).

They conclude that the given approach performed better than the benchmark approaches, especially when the number of training documents was small [2].

R. J. Mooney et al. described a content-based book recommending system which improves access to relevant information and product by giving suggestion on the previous example of users like or dislike. Most of the existing recommender system do not recommend based on the users like or dislike. Most existing recommender systems use collaborative filtering methods that base recommendations on other users preferences. By contrast, content-based methods use information about an item itself to make suggestions. This approach has the advantage of being able to recommend previously unrated items to users with unique interests and to provide explanations for its recommendations. They have described a content-based book recommending system that utilizes information extraction and a machine-learning algorithm for text categorization [3].

F. Sebastiani has described the automated categorization (or classification) of texts into predefined categories has witnessed a booming interest in the last 10 years, due to the increased availability of documents in digital form and the ensuing need to organize them. In the research community the dominant approach to this problem is based on machine learning techniques: a general inductive process automatically builds a classifier by learning, from a set of pre-classified documents, the characteristics of the categories. The advantages of this approach over the knowledge engineering approach (consisting in the manual definition of a classifier by domain experts) are a very good effectiveness, considerable savings in terms of expert labor power, and straightforward portability to different domains. He has discussed the main approaches to text categorization that fall within the machine learning paradigm. He has given in detail issues pertaining to three different problems, namely, document representation, classifier construction, and classifier evaluation [4].

Sarah Zelikovitz et al. have described a method for improvement of short text classification using a combination of labeled training data and secondary unlabeled documents. They have presented a method to reduce error rates in text classification by using a large body of potentially uncoordinated background knowledge. They used WHIRL tool which is an information integration tool that is specially designed to query and integrate varied textual sources from the Web. They have tested their system on four distinct text-categorization task that have taken from the World Wide Web. In these all of four tasks substantial reduction in error rate particularly when the set of labeled example was small. But efficiency of queries will be an issue to this work [5].

B. Sriram et al. proposed an approach to classify tweets into general but important categories by using author information and features within the tweets. This system provides ability to user to view certain types of tweets based on their interest. Traditional classification methods such as Bag-Of-Words have limitations because short texts do not provide sufficient word occurrences. They addressed this problem by using a small set of domain-specific features extracted from the authors profile and text. So that their approach effectively classifies the text to a predefined set of generic classes such as News, Opinions, Event, Private Messages and Deals. They extracted 8 feature (8f) which consist of one nominal (author) and seven binary features. Experiments are conducted with the available implementation of Naive Bayes classifier in WEKA using 5-fold cross validation. Experimental results show that BOW approach performs decently but 8F performs significantly better with this set of generic classes [6].

III. SYSTEM ARCHITECTURE

The architecture of OSN services is three-tier architecture (Figure 1)

These 3 layers are

A) Social Network Manager:- It provide the basic OSN Functionalities i.e. profile and relationship management.

B) Social Network Applications:- It provides external support for Social Network Applications. The core components of this architecture are

Short Text Classifier (STC):- Used to classify messages according to a set of categories.

Content-Based Messages Filtering (CBMF):- It exploits the message categorization provided by the STC module to enforce the FRs specified by the user.

C) Graphical User Interfaces:- It provide user interaction with the system to setup and control filtering rules and blacklist. GUI gives filtering wall where only authorized user can post.

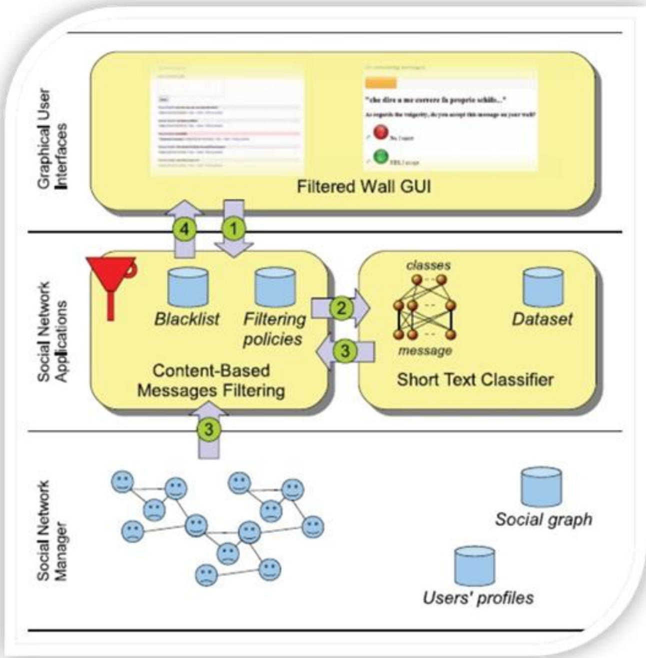

Fig 1: System Architecture [1]

As shown in Fig1, the path followed by a message, from its writing to the possible final publication can be summarized as follows:

1) Once the user tried to post message on the wall of his/her friend, then message/post is interrupted by FW.

2) A ML-based text classifier extracts metadata from the content of the message and provide to the FW.

3) FW uses metadata provided by the classifier with the data extracted from the social graph and user profiles, to enforce the filtering and BL rules.

4) Depending on the result of the previous step, the message will be published or filtered by FW.

The core components of our system are Short text Classifier (STC) and Content-based Message Filtering (CBMF).

IV. TEXT CLASSIFIER

Text classification sorts documents into a fixed number of predefined categories. Handling a large number of documents can become complicated. Thus, a text classifier places these documents into groups which are relevant to their content and makes it easier to sort them when a search for a specific document is carried out. Existing techniques used for text classification work well on large dataset such as newswires, corpora[7], but does not very well when the document in the corpus are short. So we analyze two methods for text classification, first is Machine learning based classifier Radial Basis Function Network and second is Naive Bayes classifier. Generally classification can be done in 4 steps

1) Preprocessing

2) Indexing

3) Feature Selection

4) Apply Classifier

In Preprocessing stop words and special words are removed from document, because this word does not have importance in the text categorization.

4.1 Radial Basis Function Network

Radial Basis Function Network is a type of Artificial Neural Network for supervised learning. It uses RBF as a function which is usually Gaussian and the outputs are inversely proportional to the distance from the center of the neuron.

RBFN typically have three layers: an input layer, a hidden layer with a non-linear activation function and a linear output layer. RBFN's has hidden layer of processing unit with local activation function as Gaussian Function:

$$\varphi_j = \exp\left(-\frac{\sum_{i=1}^{n}(x_i - w_{ij})^2}{2\sigma^2}\right) \quad (1)$$

All input are connected to each hidden neuron, then size or length of each vector can be calculated by Euclidian distance:

$$r_j = \sqrt{\sum_{i=1}^{n}(x_i - w_{ij})^2} \quad (2)$$

At the input layer, each input is scaled by input weight which represents weight connection between the input unit to the hidden unit. The input layer is just a fan-out layer, do not perform any processing. The Second layer performs non-linear mapping from input space into higher dimensional space in which pattern become linearly separable. The output of the hidden unit is calculated by Euclidian distance. The third layer performs a simple weight sum from hidden unit [8].

The Short text categorization is a hierarchical two-level classification process. The first-level classifier does a binary classification that labels messages as Neutral and Non-Neutral. The first-level filtering task enables the subsequent second level task in which a finer-grained classification is performed. The second-level classifier carries out a soft-partition of Non-neutral messages assigning a given message a gradual membership to each of the non-neutral classes. Such a list of grades is then used by the subsequent phases of the filtering process.

In the case of filtering of unwanted message from OSN user wall, preprocessing can be done first of all, after that RBFN can be applied to preprocessed data so that we get categorized data. This categorized data can be used by filtering rule to remove unwanted message.

4.2 Naive Bayes

Naive Bayes classifiers are a family of simple probabilistic classifier. It is based on Bayes theorem with strong independence assumption between features. It uses Bayes theorem to calculate the probability of the message whether it is neutral or non-neutral. Each word has probability of occurring in neutral message (good messages) or non-neutral message (bad messages). Initially system doesn't know probability of that word. So it must trained to calculate probability of each word. So in the trained set, user must manually indicate whether new message from which category. For all word in each training message this classifier adjust the probability of each word will appear in which category in database. We want to know given massage x is from which category c, that is we want to know P(C|X) and which is can be computed by using Bayes rule [9] as:

$$P(c/x) = P(x/c) * P(c) / P(x) \tag{3}$$

P (c/x) is posterior probability of class (target) given predictor (attribute).
P (x/c) is likelihood which is the probability of predictor given class.
P (c) is prior probability of the class.
P (x) is prior probability of predictor.
In this way message probability is computed over all the word, if it exceeds certain threshold then filtering will mark message as neutral and non-neutral in first level. After that non-neutral message are sub categorized into five categories as Violence, Vulgar, Offensive, Hate and Sex.

In the case of filtering of unwanted message from OSN user wall, preprocessing can be done first of all, after that Naive Bayes classifier can be applied to preprocessed data so that we get categorized data. This categorized data can be used by filtering rule to remove unwanted message.

V. MANAGEMENT OF FILTERING RULES AND BLACKLIST

We model social network as a directed graph in which each node represent to network user and link represent relationship between them.

5.1 Filtering Rules

The same message on OSNs may have different meanings and relevance based on who writes it. It is necessary to apply constraints on messages. Constraints can be selected on several different criteria's. User can state what contents should be blocked or displayed on filtered wall by means of Filtering rules. Filtering rules are specified on the basis of users profile as well as user social relationship. FR is dependent on following factors.

1) Author

2) Creator Spec

3) Content Spec

4) Action

An author is a person who defines the rules. creatorSpec denotes the set of OSN user and their relationship and contentSpec is a boolean expression defined on content in form of (C,ml), where C is class of the first or second level and ml is the threshold of minimum membership level required for class C to make constraint satisfied. Action denotes the action to be performed by the system on the message whether publish, delete, shuffle or null. In the case of neutral message we used action as publish. In the case of vulgar, offensive and sex we used to action as delete, in the case of hate we have used action as shuffle and in case of violence we have used action as null. Filtering rule FR can be in the form of tuple as (author, creatorSpec, contentSpec, action). We have used following filtering rule to provide restriction to unwanted messages displaying on user walls. Some of filtering rules examples are:

Example 1 of filtering Rule:

Bob is user, and he want to block the user who is in his friendliest, if he tries to post vulgar a message having threshold of certain amount then filtering criteria can be easily specified as

(Bob, friendof,(Vulgar,0.80),delete)

If friend of Bob tried to post "U are stupid" then if suppose stupid is vulgar word who have threshold 0.80 then the whole message can't display on the wall.

But if message contains important data but having few hate or violence word then if user want to display message on wall so we add new actions like nulling, or shuffling. Examples of these are as follows.

Example 2

(Bob, friendof,(Violence,0.80),null)

If friend of Bob tried to post "U are stupid" then if suppose stupid is Violence words who have threshold 0.80 then only that violence word are getting null and remaining message is display as it is. In this case it display message as "U are null".

Example 3

(Bob, friendof,(Hate,0.80),shuffle)

If friend of Bob tried to post "U are stupid" then if suppose stupid is Hate word who have threshold 0.80 then only that hate words are getting shuffle and remaining message is display as it is. In this case it display message as "U are tuipsd".

5.2 Blacklist

BL users are those users whose messages are prevented independent from their contents. BL rules enable the wall owner to determine users to be blocked on the basis of their profiles and relationship with wall owner. This banning can be done for a specified period or forever according wall owner's desire. Like FR, BL is also dependent on author, creator specification and creator behavior. For denoting users bad behavior, we have considered this measure. For the blacklist management, we have considered Relative Frequency (RF), it detect those users whose messages continue to fail the FRs. If threshold or relative frequency is greater than 80%, then the user directly inserted into blacklist and it can't post message on wall until it is unblocked by any of his friend.

VI. EXPRIMENTAL RESULTS

6.1 Dataset

We have used available WmSnSec dataset. WmSnSec dataset consist of 1266 messages from publically accessible Italian groups. It consist of message body and the name of class to which it belongs. The set of classes considered Ω={neutral, violence, vulgar, offensive, hate, sex} where Ω -neutral are the second level classes [10].

6.2 Result Evaluation

We have measured the time, precision, recall and F-measure to evaluate the effectiveness of both classifiers which are shown in Table I.

Table I-Time, Precision (P) ,Recall (R) , F-measure(F1) value for RBFN and Naive Bayes with Filtering Rules

Classifier	Time	Precision	Recall	F-measure
RBFN	65332ms	80%	67%	73%
Naïve Bayes	39418ms	78%	64%	70%

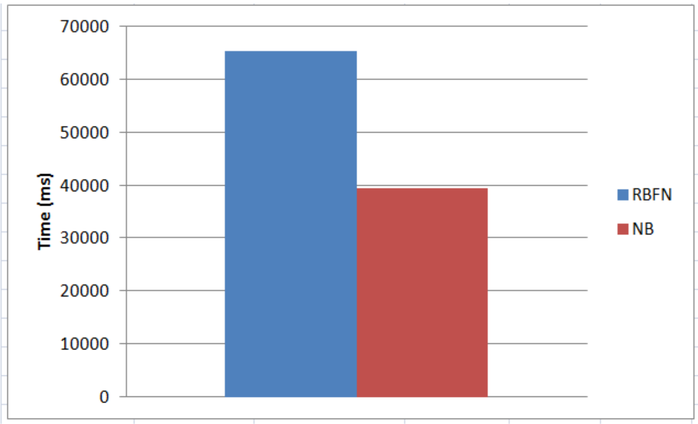

Fig 2: Comparison of RBFN and Naive Bayes in terms of Learning Time for WmSnSec Dataset

Table I shows that, RBFN classifying filtering approach is better to classify and categorize text but it take more learning time as comparative to Naive Bayes classifier.

VII. CONCLUSION

This paper presents the comparative performance evaluation of Radial basis function network and Naive Bayes for text classification and design of system to filter undesired messages from Online Social Networking user wall. Additionally the flexibility of a system can be enhanced through filtering rules and blacklist management. We have used WmSnSec dataset and after evaluating performance we conclude that RBFN classifying filtering approach is better, but it take more learning time comparative to Naive Bayes classifier.

REFERENCES

[1] 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, Feb 2013.

[2] 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.

[3] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," *in Proceedings of the Fifth ACM Conference on Digital Libraries*. New York: ACM Press, 2000, pp. 195–204.

[4] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[5] Zelikovitz and H. Hirsh, "Improving short text classification using unlabeled background knowledge," *in Proceedings of 17th International Conference on Machine Learning (ICML-00)*, P. Langley, Ed. Stanford, US: Morgan Kaufmann Publishers, San Francisco, US, 2000, pp. 1183–1190.

[6] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," *in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR* 2010, 2010, pp. 841–842.

[7] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, "Rcv1: A new benchmark collection for text categorization research," Journal of Machine Learning Research, 2004.

[8] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," Neural Computation, vol. 1, p. 281, 1989.

[9] Saurabh Kr. Srivastava1 and Sandeep Kr. Singh "Multi-Parameter Based Performance Evaluation of Classification Algorithms" in proceeding of International Journal of Computer Science Information Technology (IJCSIT), Vol 7, No 3, June 2015.

[10] WmSnSec dataset available at http://www.marcovanetti.com/pages/wmsnsec.