ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

Content Based Message Filtering System for OSN User Walls

Neethu S[1], Rajesh Kumar B[2]

P.G. Scholar, Department of CSE, R.V.S. College of Engineering and Technology, Coimbatore, India[1].

Assistant Professor, Department of CSE, R.V.S. College of Engineering and Technology, Coimbatore, India[2].

ABSTRACT: Online Social Network plays a vital role in our each day life. A person can communicate with others by sharing information in several forms like image, audio and video contents. Major matter in OSN (Online Social Network) is to provide security in posting unnecessary messages. One fundamental question in present Online Social Networks (OSNs) is to give users the capability to have power over the messages posted on their individual private space to avoid that unnecessary content is displayed. Till today , OSNs contribute slight support to this need. As a solution to this problem in this paper, we propose a system in which OSN users to have a straight control on the contents displayed on their private space. This is done with the support of a flexible constrain-depended system, that permits users to provide the filtering to be applied to their walls, and a Machine Learning-based classifier automatically labelling messages in support of content-based filtering

I. INTRODUCTION

Today's modern life is totally based on internet. now a days people cannot imagine life without internet. also, osns are just a part of modern life. From last few years people share their views, ideas, information with each other using social networking sites. Such communications may involve different types of contents like text, image, audio and video data. but, in today's osn , there is a 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.

Osnsgive support to filter unwanted messages on user walls. for instance, facebookpermits users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends). However, no content-based preferences are supported and thusit'sunfeasibleto controlunsought messages, like political or vulgar ones, notwithstanding of the user's identity who posts them. this is oftenas a result of wall messages areofficial by short text thatancient classification ways have serious limitations since short texts don'tgivespare word occurrences.

The aim of this work is thus to propose and by experimentation valuate an automatic system, referred to as Filtered Wall (FW), ready to filter unwanted messages from OSN user walls. we tend to exploit Machine Learning (ML) text categorization techniques to mechanically assign with every short text message a collection of classes supported its content the main efforts in building a sturdy short text classifier (STC) square measure targeted within the extraction and choice of a collection of characterizing and discriminant options. The solutions investigated in this paper is an easy way from that we tend to inherit the training model and also the induction procedure for generating pre classified knowledge. The initial set of options, derived from endogenous properties of short texts, is enlarged here as well as exogenous information associated with the context from that the messages originate. As way because the learning model is bothered, we tend to ensure within the current paper the utilization of neural learning that is these days recognized mutually of the foremost economical solutions in text classification.

we base the short text classification strategy on back propagation which is a neural network technique and is having established capabilities in acting as soft classifiers, in managing shrie information and in and of itself imprecise categories. Moreover, the speed in performing the training section creates the premise for associate adequate use in OSN domains, additionally as facilitates the experimental analysis tasks. A back propagation network is a man-made neural network that uses a neural functions as activation functions. The output of the network may be a linear combination activation functions of

10.15680/ijircce.2015.0302011

the inputs and vegetative cell parameters. It perform networks have several uses, together with perform approximation, statistic prediction, classification, and system management.

The speed in activity the training part creates the premise for Associate in Nursing adequate use in OSN domains, also as facilitates the experimental analysis tasks. Back propagation functions area unit the native tuned process units. This network can have few layer of units with a selective response for a few vary of the input variables. every unit has Associate in Nursing overall response operate, presumably a Gaussian: especially, the general short text classification strategy on back propagation operate Networks is employed for its verified capabilities in acting as soft classifiers, in managing clattering information and as such imprecise categories. A Back propagation network is a man-made neural network that uses Back propagation functions as activation functions. The output of the network could be a linear combination ofback propagation functions of the inputs and nerve cell parameters. Back propagation operate networks have several uses, together with operate approximation, statistic prediction, classification, and system management. Moreover, the speed in activity the training part creates the premise for Associate in Nursing adequate use in OSN domains, also as facilitates the experimental analysis and classification.

II. RELATED WORK

Content-based filtering is especially supported the employment of the metric capacity unit paradigm consistent with that a classifier is mechanically iatrogenic by learning from a group of preclassified examples. a noteworthy sort of connected work has recently appeared, that dissent for the adopted feature extraction strategies, model learning, and assortment of samples. The feature extraction procedure maps text into a compact illustration of its content and is uniformly applied to coaching and generalization phases. many experiments prove that Bag-of-Words (BoW) approaches yield sensible performance and prevail normally over additional subtle text illustration that will have superior linguistics however lower applied mathematics quality

As so much because the learning model cares, there square measure variety of major approaches in content-based filtering and text classification normally showing mutual blessings and downsides in perform of application dependent problems. Boosting-based classifiers , Neural Networks, and Support Vector Machines over different in style strategies, like Rocchio and Naı¨ve Bayesian . However, it's price to notice that the majority of the work associated with text filtering by metric capacity unit has been applied for long-form text and also the assessed performance of the text classification strategies strictly depends on the character of matter documents

III. ALGORITHM

The overall short text classification strategy on back propagation networks A back propagation network is a synthetic neural network that uses back propagation functions as activation functions. The output of the network could be a linear combination of back propagation functions of the inputs and vegetative cell parameters. back propagation networks have several uses, as well as perform approximation, statistic prediction, classification, and system management. Moreover, the speed in playacting the educational section creates the premise for associate degree adequate use in osn domains, still as facilitates the experimental analysis tasksback propagation functions square measure the native tuned process units. this network can have few layer of units with a selective response for a few vary of the input variables. every unit has associate degree overall response perform, probably a gaussian: particularly, the short text classification strategy on back propagation networks is employed for its verified capabilities in acting as soft classifiers, in managing vociferous knowledge and per se obscure categories. a back propagation network is a synthetic neural network that uses radial basis functions as activation functions. the output of the network could be a linear combination of activation functions of the inputs and vegetative cell parameters. back propagation networks have several uses, as well as perform approximation, statistic prediction, classification, and system management. moreover, the speed in playacting the educational section creates the premise for associate degree adequate use in osn domains, still as facilitates the experimental analysis tasks.

- Pre-Processing
- Content based Filtering
- Short text classifier and Content text representation

- Managing blacklist and filtering rules
- Performance evaluation

Step 1: Pre-Processing
Technology of stop words removal for deleting meaningless words was used. Stemming algorithm was also applied to remove the affixes (prefixes and suffixes) in a word in order to generate its root word. The stop words are the words that frequently occur in documents. Eliminating the stop words in automatic indexing speeds the system processing, saves a huge amount of space in index, and does not damage the retrieval effectiveness. There are various approaches used for determination of such stop words. Nowadays, there are several English stop words lists that are commonly used in information retrieval. Proposed method was carried out to eliminate all the stop words in the text to speed up the system processing.

Stemming is a solution for one of the problems involved in information retrieval, like variation in word forms. The most common types of variation are spelling errors, alternative spelling, multiword construction, transliteration, affixes and abbreviations. These variations in words form lead to the efficiency issue in the matching algorithm during the information retrieval process. Using root word in pattern matching provides a much better effectiveness in information retrieval.

Step 2: Content based Filtering
In content based filtering to ascertain the user's interest and former activity in addition as item uses by users best match is found. as an example OSNs like Facebook, Orkut used content based mostly filtering policy. in this by checking users profile attributes like education, work area, hobbies etc. recommended friend request might send. the most purpose of content based mostly filtering, the system is in a position to find out from user's actions associated with a selected content supply and use them for different content sorts.

In content-based filtering every user is assumed to control severally. As a result, a content-based filtering system selects data things supported the correlation between the content of post and also the user preferences as opposition a cooperative filtering system that chooses items supported the correlation between folks with similar preferences.Documents processed in content-based filtering area unit largely matter in nature and this makes content-based filtering near text classification.

Step 3:Short text classifier and Content text representation
In our approach the short text classifier is named as multi class soft classification method. The back propagation categorizes short messages as Neutral and Non-neutral; within the second stage, Non-neutral messages are further classified into subclasses like vulgar , political ,etc. Extracting the actual options from a given set of knowledge is crucial task; this may have an effect on the whole performance feature of classification. Text illustration is named as vector model within which text document is diagrammatic as dj and also the real weight of the dj=$w1j$,….,, t is that the set of term that occur a minimum of once from the gathering of documents Tr and $wkj\epsilon$ [0, 1] represent tk contribute to document dj.in back of words. just in case of non-binary weight wkj, document dj is computed in line with the Inverse Document Frequency weight perform is outlined as

$$tf{-}idf\ (tk,\ dj) = \#(tk,dj).\log \frac{|\,\mathrm{Tr}\,|}{\#\,\mathrm{Tr}\ (\mathrm{tk})}$$

Where # (tk, dj) denotes the quantity of times tk occur in document dj and #Tr (tk) denote document frequency of tk.

Step 4: Managing blacklist and filtering rules
Blacklist rule is used to avoid unnecessary post or content created by the unapproved users. This is the tool which is managed by the system, such that the system may able to decide who are users enter into the blacklist and how long the message can be present in the system. That is the wall owner must decide who are the users enter into his/her private wall and how long they present in the wall. The wall owner must specify the blacklist rules. This blacklist management will block the users based on the user profiles and the relationship in the online social networks such that the wall owner key entity to identify the users.

Filtering rule is said as the writer or maker who specifies the rules take an example assume jijo is an online social network user and he continuously wishes to block the maximum political messages. And jijo want to screen only the message coming from secondary friends and not for the straight friends. This can be done using filtering rule as follows:

((rajev, friendof, 2,0. 1), (political, 0.50), block)
((rajev, friend of, 1, 0.3), (political,0.60), block)

The values given in thecase are trust value 0.60 and 0.50.if the friend of jijo with the trust value of 0.60 wants to publish the message as “political violence in india”on jijo private wall. Afterward placement of this message it will create the rating value as 0.65 for the class political message. So, the message that holding massive degrees of political content will be cleared from the system and the filtered message will not show in the user isolated wall.

It decreases the undesirable content and expands the performance of the system. Moreover, filtering rule is the action done by the system on the message to fulfill the rule. It hang onthe wall owner’s likings and also the classification outcomes.

Step6: Performance analysis

The existing and also the planned protocols performance area unit evaluated.during this module performance analysis is finished supported the Recall, Precision. The performance of the system isn't compared with the other system however merely evaluated the performance of the filtering system in line with the rise within the coaching files. The systems potency to supply the proper result's examined against the rise within the coaching files. A graph is premeditated by taking prediction in y axis and no of coaching files in x axis.

IV. RESULTS AND DISCUSSION

In order to supply Associate in Nursing overall assessment of however effectively the system applies a Fr, In distinction, Recall needs to be understood because the likelihood that, given a rule that has got to be applied over an explicit message, the rule is actually enforced . the exactness and also the Recall worth computed for FRs with ðNeutral; 0:5Þ content constraint. In distinction, the unit stores the exactness and also the Recall worth computed for FRs with (V ulgar; 0:5) constraint

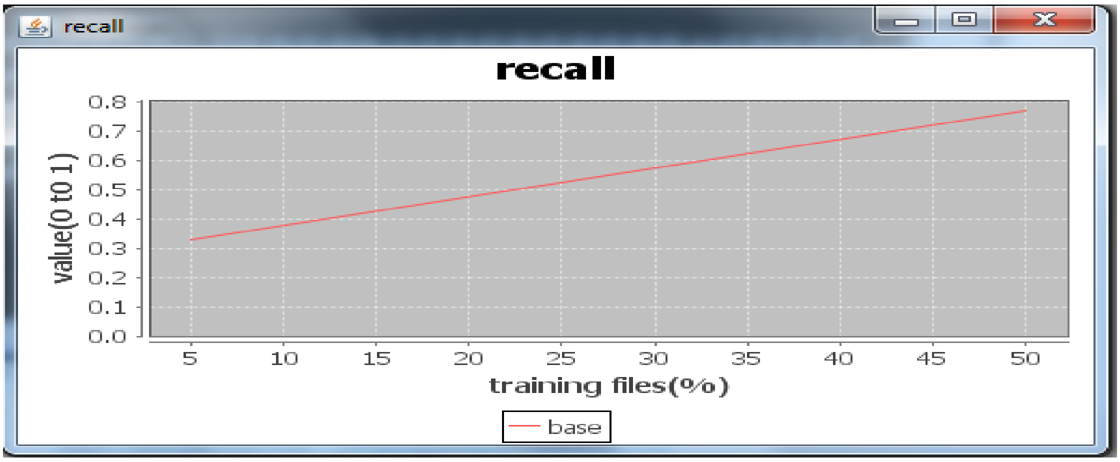

Fig.1 performance evaluation using recall

The existing and also the planned protocols performance area unit evaluated. During this module performance analysis is finished by using recall.

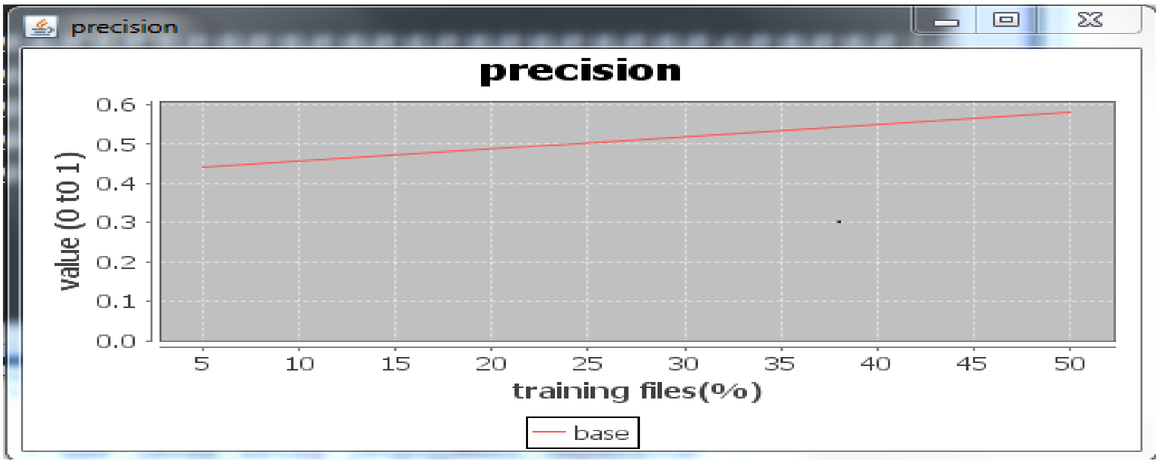

Fig.1 performance evaluation using precision

Results achieved by the content-based specification part, on the first-level classification, is thought-about good enough and fairly aligned with those obtained by well-known info filtering techniques.Results obtained for the content-based specification part on the second level area unit slightly less good than those obtained for the primary, however we must always interpret this visible of the intrinsic difficulties in distribution to a messages a semantically most specific class (see the discussion in Section six.2.2). However, the analysis of the options rumored in Table one shows that the introduction of discourse info (CF) considerably improves the flexibility of the classifier to properly distinguish between non neutral categories. This result makes a lot of reliable all policies exploiting non neutral categories, that area unit the bulk in real-world situations

V. Conclusion and Future Work

A system to filter unwanted message in OSN wall is given. the primary step is to classify the content victimization many rule. Next step is to filter the unwanted posts or content. Finally Blacklist rule is enforced. so owner of the user will insert the user who post unwanted messages. higher privacy is given to the OSN wall holder in our system. This work is that the beginning of a wider project. the first encouraging results we've obtained on the classification procedure prompt us to continue with alternative work which will aim to boost the standard of classification. especially, future plans ponder a deeper investigation on 2 mutualist tasks. the primary considerations the extraction and/ or choice of discourse options that are shown to own a high discriminative power. The second task involves the educational section. Since the underlying domain is dynamically dynamic , the gathering of pre-classified knowledge might not be representative within the long run.

In future Work, we have a tendency to attempt to implement the filtering rules that stop the actions with the aim of bypassing the filtering system. It is enforced by victimization the tactic of finite automata.

REFERENCES

1. 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.
2. Brian Archibald Cmu , George Doddington , Jaime Carbonell , James Allan, George Doddington , Jonathan Yamron , Yiming Yang, “Topic Detection and Tracking Pilot Study Final Report”, 1998
3. ChengXiang zhai, Qiaozhu Mei, “Discovering Evolutionary Theme Patterns from Text: An Exploration of Temporal Text Mining” Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, 198-207, 2005.
4. Christopher S.G, Jin- Cheon Na, , Subbaraj Shakthikumar, Tun Thura Thet ,“Sentiment analysis of movie reviews on discussion boards using a linguistic approach”, Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, 2009
5. Jantce Wiebe, Paul Hoffmann, Theresa Wilson, “Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis”, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005

6. John M. Trenkle, William B. Cavnar , "N-gram based text categorization", In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval, 1994
7. Mike Thelwall, Rudy Prabowo, "Sentiment Analysis: A Combined Approach", Journal of Informetrics, Volume 3, Issue 2, April 2009, Pages 143–157

BIOGRAPHY

Neethu.S is pursuing Master of Engineering in Computer Science Engineering in R.V.S College of Engineering and Technology, Coimbatore, India.

B.Rajesh Kumar is an Assistant Professor in Department of Computer Science,R.V.S College of Engineering and Technology,Coimbatore,India.