ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)
Vol. 2, Issue 2, Ver. 1 (April - June 2014)
International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)

A Machine Learning Approach for Filtering Unwanted Comments Posted in Online Social Networks

[I]G.Parvathi, [II]N.Sakthidevi, [III]Veena Alphonsa Jose
[I,II]M.E Student, CSE, Srinivasan Engineering College
[III]Assistant Professor, CSE, Srinivasan Engineering College

Abstract
The main feature in online social network is any one can able to post any kind of messages to any user wall. The fundamental issues in online social networks (OSN) are preventing unwanted comments which are posted in OSN user's private wall. The comments posted in online social network do not have proper word occurrences. That comment consists of short text and non-English word. To overcome this obstacle, many traditional classification methods are used to classify the short text and non-English word but still it is need to be improved. So, by proposing machine learning approach, unwanted comments are filtered out. This paper presents a general framework for building the classifiers that deals with short text and non-English words. The OSN users who one is trying to posting unwanted comments are blocked by using flexible rule based system.

Index Terms
Online Social Network, Short text classification, filtering rules, machine learning approach, Information retrieval.

I.Introduction

This paper is under the domain of data mining. Data mining refers generally to the process of extracting interesting information and knowledge from unstructured text. The main goal of text mining is to enable users to extract information from textual resources. The most important part of text mining is the text categorization. Text Classification (TC), also known as text categorization or topic spotting, is the automatic classification of text documents under predefined categories. Machine learning helps to categorize the documents automatically. Information Retrieval helps to characterize the text as a feature.

Today one of the most popular interactive medium is Online Social Networks, which is used to communicate with many people to share knowledgeable information, and distribute a considerable amount of human life information. Every day communications imply the exchange of several types of information including text, image, audio and video data. Overall statistics of Face book reveals that the average user creates 90 pieces of information in every month, whereas more than 30 billion pieces of information such as news stories, blog posts, notes, web links, and photo album are shared by each month. The huge and dynamic character of the data, Creates premises of employment of online information mining strategies. Information filtering and text mining mechanism are used to avoid the overwhelmed by useless data. The step by process of information filtering is collect the data, perform pre-processing step for removing stop words, stemming process and applying machine learning approach for classification of data or text into predefined categories. Machine learning, a branch of artificial intelligence, concerns the creation and study of systems that can learn from data. For example, a machine learning system could be trained on email messages to learn to differentiate between spam and non-spam messages. It can then be used to classify new email messages into spam and non-spam folders.

The foundation of machine learning deals with representation and generalization. Representation of data instances and function evaluated on these instances are part of all machine learning systems. Generalization is the property that the system will perform well on unseen data instances. A core objective of a learner is to generalize from its experience. Generalization is the ability of a learning machine to perform correctly on new, unseen tasks after having experienced a learning data set. The training examples come from some generally unknown probability distribution. The learner has to build a general model about this space that enables it to produce adequately accurate predictions in new cases. Machine learning focuses on prediction, which is based on known properties learned from the training data. In machine learning approach, a supervised learning algorithm is enforced. A supervised learning is a class of machine learning technique that make use of both labeled and unlabeled data for training a small amount of labeled data with a large amount of unlabeled data. This technique is in between unsupervised and supervised learning approach. Semi-supervised learning produces considerable improvement in learning accuracy. A main motivation of this paper is to prevent or filter the unwanted comments posted by the OSN users. According to online social network, the comments posted in user private wall consists of a short text (e.g., wru instead of "where are you", hru instead of "how are you") and non- English words (e.g., enna panra instead of "what doing") in a Tamil language. Many existing techniques are used for short text classification, even though some kind of unwanted messages such as political, vulgar can't be classified efficiently.

The main objective of this paper is to prevent or filter the unwanted comments which are posted in Online Social Networks. The user who is tried to post unwanted comments are filtered and blocked based on the message and profile relationship of the user. This is due to the main feature that in OSNs there is the possibility of posting or commenting any other posts on particular public/ private areas. 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. Today OSNs provide very little support to prevent unwanted messages on user walls. For example, Face book allows OSNs users to state who is allowed to insert messages in their walls (i.e., friends of friends, friend or defined groups of friends or unknown). However, no content-based preferences are maintained and therefore it is not possible to prevent unwanted messages, such as political or vulgar ones, no matter of the user who posts them. The user's private wall messages are constituted by short text for which traditional

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

classification methods have serious limitations since short texts do not provide sufficient word occurrences.

II.Related Work

Many existing text classification techniques are used to classify the text documents. In general, text classification includes topic based text classification and text genre-based classification. Topic-based text categorization classifies documents according to their specific topics. Texts can also be written in many genres, for instance, news reports, scientific articles, movie reviews, and advertisements. Genre classification is the process of grouping objects together based on defined similarities such as subject, arrangement, style. Genre classification is a means of managing information is already established in music and text and is used, along with topic classification, to organize materials in the commercial sector and intellectually. However, in the case of text, genre is not a well-defined notion (it is better defined in music and arts) and discussions over what exactly constitutes genre abound in the classification community.

Text Classification

Some of the existing text classification techniques are Decision tree, Nearest Neighbor, Naïve bayes, neural networks, Rocchio`s Algorithm, Support vector machine etc. A bag-of-words technique is mostly used to classify the text document. Naive Bayes is mostly used in many text classification applications because of its effectiveness and simplicity. To improve the performance of the text classifier a multinomial approach is enforced [17]. Even though the performance is degraded because it does not classify or model text properly since the lack of attributes feature selection. So a simple transformation is used effectively. The duplicate words in a document to account for burstiness phenomena in text are removed. When documents are very short the uniform priors with skewed class distributions are used. It is performed better than the mutinomial naïve bayes approach [9].The author [10] suggests tree-like Bayesian networks which are able to handle a text classification task in one hundred thousand variables with sufficient speed and accuracy.

Support Vector Machines eliminate the need for feature selection and making the application of text classification considerably easier. Another advantage of SVMs methods is robustness. SVMs give good performance in all experiments, avoiding catastrophic failure. Support Vector Machines do not require any parameter tuning, since it can find good parameter settings automatically. This makes Support Vector Machine is a very promising and easy-to-use method for learning text classifiers. This classification provides excellent precision but poor recall. To improve a recall a generic SVM method is used with adjusting threshold [18]. A program can able to perform class separation, overlapping classes,nonlinearity,problem solution then that is called as support vector machine[4].

Matt Healy., et.al. Proposed a Case based approach [11] in which the short text messages are classify using Email Classification Using Examples (ECUE) system. It is used in email spam filtering. The EUCE System extracts three types of features. First one is word features; second one is single characters, features and statistical features. The system uses binary feature representation to produce the best generalization accuracy for spam filtering. It uses K-Nearest Neighbor (K-NN) with weighted distance voting. It also uses competence editing technique to remove noisy and redundant data.

Sarah Zelikovitz., et.al. Proposed a transtuctive Latent Semantic Indexing [13] is based upon the assumption that there is an underlying semantic structure in textual data, and that the connection between terms and documents can be described in this semantic structure form. Textual documents are represented as vectors in a vector space. In a vector a term represents position i equal to 1 if that term appear in the document and equal to 0 if the term does not appear in the document. Background knowledge concerns cases where the data to which the learned classifier will be applied is available at the start of learning. For such learning problems, called transductive learning these unlabeled examples may also prove helpful in improving the results of learning. Latent Semantic Indexing (LSI) tries to overcome the problems of lexical matching by using statistically derived conceptual indices instead of individual words for retrieval.

Joachims.T.et.al. proposed a Rocchio's Algorithm[6] is based on the Relevancy Feedback Algorithms. These kinds of Relevancy Feedback Models are an effective way of modifying and expanding user queries. Rocchio's Algorithm is one of the earliest methods used for queries. This algorithm is based on the idea that if the relevance for a query is known, an optimal query vector will increase the average query-document similarity for relevant documents, and will simultaneously decrease query-document similarity for non relevant documents. Rocchio's Algorithm has a poor performance when the proportion of relevant documents in the whole corpus is low.

Bobicev.V. et.al. Proposed a Prediction by partial matching (PPM) [2] , is an adaptive finite-context method for text compression. That is a back-off smoothing technique for finite-order Markov models. It obtains all information from original data, without feature selection. So it is easy to implement and relatively fast. The character-based Prediction by Partial Matching models were used for spam detection, source-based text classification and classification of multi-modal data streams that included texts. Treating a text as a string of characters, a character-based Prediction by Partial Matching avoids defining word boundaries. Prediction by partial matching deals with different types of documents in a uniform way. Noisy texts and contain errors of different types that affect the word-based methods. Letter-based methods avoid noisy text and errors and it captures the characteristics of the text. The letter-based model is more compact and faster to build than the word based methods.

III. Techniques Used

Supervised Classification

Supervised classification consists of inducing hypothesis (h) of an unknown function f is defined from an input space X to a discrete unordered output space Y=f1;;;;;kn, given a training set S. The training set contains K training examples, where s=f(x_1,y_1);;;;;;(x_k,y_k)n, which are pairs of (x,y) where x belongs to X and y=f(X).The values of output space Y is associated with the each training examples are called as class.Therefore,training examples completely describe a set of attribute-value pairs and a class label. In machine learning, the term supervised Learning refers to the fact that the training examples are interpreted with the class labels, which are taken from a pre-classified set. To training a data set this approach is used.

To improve the classification based on the filtering concepts the sense or semantic of the word is more important. When training a set, word sense disambiguation(WSD) also have to consider

improving classification. It is mainly consider a set of words co-occurring in a document will determine the proper senses for one another word even though each individual word being multiplied ambiguous. Example WordNet tool. WordNet is a lexical database for the English language. It grouping English words into sets of synonyms called synsets. It provides general definitions, records and short the various semantic relation between these synsets. It is mostly used to support automatic text analysis and artificial intelligence applications.

It also provides the polysemy count of a word.That mean the number of synsets that contain the word. If a word contributes in several synsets (i.e. has several senses) then typically some senses are much more common than others. WordNet measures this by the frequency score: in which several sample texts have all words semantically tagged with the matching synset, and then a count provided representing how often a word appears in a specific sense. WordNet is mostly used in Word sense disambiguation (WSD).

Short text classification

Classifying short text is very difficult process. In short text classification the most important part is text representation. However, the most appropriate feature representation and feature set for short text messages have not yet been adequately investigated. To represents text as a feature, three types of features are available. They are Document properties (Dp), bag-of-words (Bow), contextual feature(CF).Here CF modeling information is used that Characterize the environment where the user is posting. These features play a key task in deterministically accepting the semantics of the messages. In the BoW representation, terms are identified with words. In the case of non binary weighting, term tk in document dj and the weight wkj of computed according to the standard term frequency-inverse document frequency (tf-idf) weighting function.

Term Frequency:

$$tf(t,d) = 0.5 + \frac{0.5 \ X \ f(t,d)}{\max\{f(w,d):w \in d\}}$$

Inverse Term Frequency:

$$itf(t,d) = \log \frac{|D|}{|\{d \in D : t \in D\}|}$$

Term Frequency – inverse term frequency:

$$tf\text{-}itf(t,d,D) = tf(t,d) \ X \ itf(t,D)$$

IV. System Design

The proposed system architecture depends on Machine Learning Approach. The design is shown in Figure 1 where the users are register into the face book account. While registration the information about the user is stored into the database. Then the user tried post comment on the user's private wall. If user sent any unwanted comments that are filtered using general framework called filter wall.

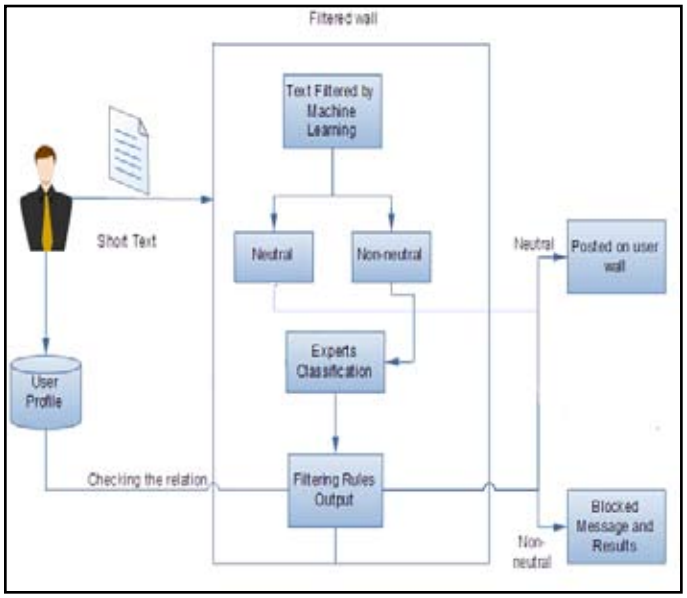


Fig.1: System Architecture

When the user tries to send a comment on another user private wall, the unwanted comments are filtered using a filter wall. Text is filtered using machine learning approach. Text which is posted by user is classified into neutral and non-neutral message. Neutral message is a good message or a ordinary message. Non-neutral message is a bad message or a unwanted or undesired message. The main objective is to filter the unwanted message being posted in the user's private wall. And the text is consists of hint's and non-english word. To filter this kind of message a distributional term representation and latent semantic indexing are used. Based on the profile information and content of the message the user is blocked. For each and every undesired message some probability value is provided. After filtering the message will not been shown on user wall. The probability value only shown. In profile checking ,it check the user is a friend or friend of friend or unknown .if he is a friend or friend of friend instead of blocking it send an alert message to indicate that user get a alert a undesired message from them with him or her name. if they want to continue to see they can see else they ignore a message. Suppose he is a unknown then the message is filtered and the user account is blocked.

V. Proposed Work

In machine learning, a supervised algorithm is used to filter the text or comments efficiently. After feature selection and transformation the documents can be easily represented in a form that can be used by a Machine Learning algorithm. Many text classifiers have been proposed using machine learning techniques, probabilistic models, etc. while many approaches have been proposed, but automated text classification is still a major area of research primarily because the efficiency of present automated text classifiers is not faultless and still needs development. Recently in the area of Machine Learning the concept of combining classifiers is proposed as a new direction for the improvement of the performance of individual classifiers. In the context of combining multiple classifiers for text classification, a number of researchers have shown that combining different classifiers can advance classification accuracy.In proposed system the unwanted comments posted by the osn users are classiffied using machine learning approach.

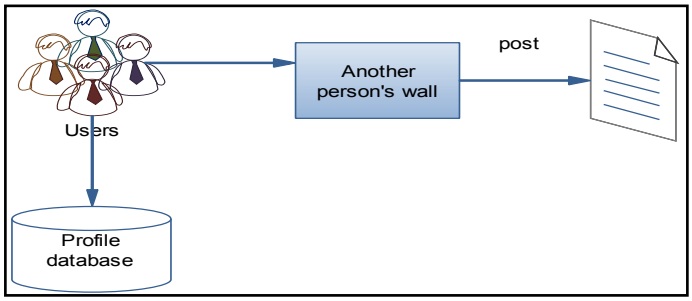
It is divided into four steps.

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

1. User Registration.
2.Content based message filtering.
3.Short text classification.
4.Filtering rules.

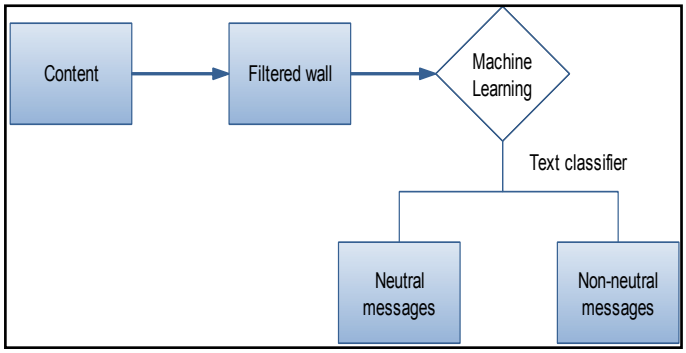
A. User Registration

User Registration and Posting Messages on User Wall. Users of the social network give their profile details to the network at the creation time. So every user gives their name, gender, profession, mobile number etc. In this system, the user's profile details in the database. After registration process user connect with their friends by friend request. Some User may want to post unwanted messages on another user's wall.



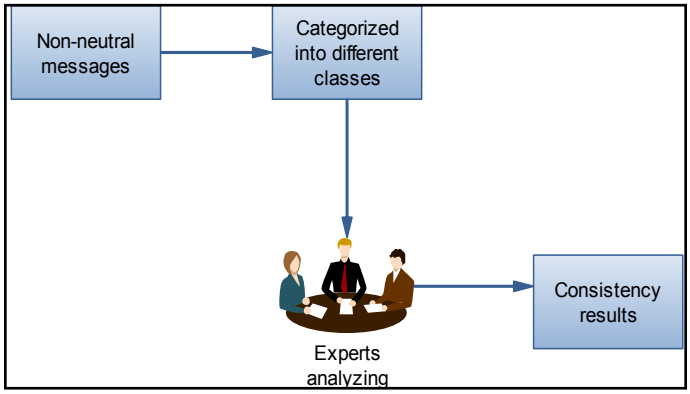
B. Content Based Message Filtering

The Filtered wall in the system to filtering the unwanted messages. Machine Learning is the text classifier which is used to classify the messages and short texts posted by users. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. Machine learning techniques classify the message into neutral and non-neutral. After that non-neutral messages are further divided into several classes.



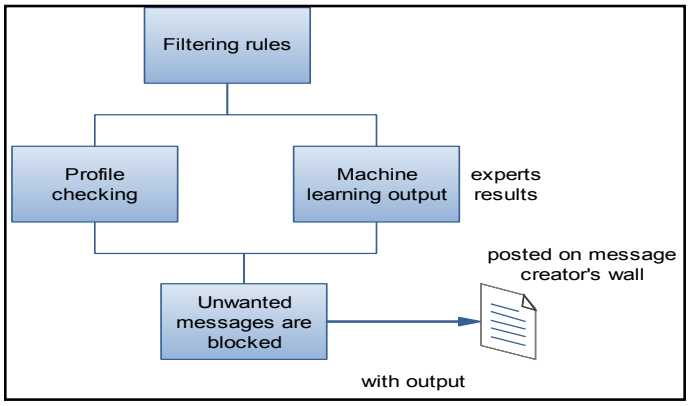
C. Short text classification

To analyze the results of the Machine Learning technique, the messages are evaluated by probability calculation. It evaluates the non-neutral messages under Violence, politics, offensive, hate, sexual harassment categories. Bag of Words, contextual features yield good performance in text categorization. It improves overall accuracy and consistency in analyzing messages and stored the results.



D. Filtering Rules

User's each and every details are stored in their profile database. Before posting the message filtering rules checking the relation between the message creator and receiver and also the profile creation date etc. In profile checking ,it check the user is a friend or friend of friend or unknown .if he is a friend or friend of friend instead of blocking it send an alert message to indicate that user get a alert a undesired message from them with him or her name. If they want to continue to see they can see else they ignore a message. Suppose he is a unknown then the message is filtered and the user account is blocked.



VI. Implementation

AdaBoost Algorithm:

Procedure AdaBoost.MH
Input: S = f(xi; Yi); 8i = 1::mg
S is the set of training examples
Initialize distribution D1
m is the number of training examples
k is the number of classes
D1(i; l) = 1
mk; 8i = 1::m; 8l = 1::k
Perform T rounds:
for t := 1 to T do
Get the weak hypothesis ht : X £ Y ! <
ht = WeakLearner(X;Dt)
Update distribution Dt
Dt+1(i; l) = Dt(i;l)exp(¡Yi[l]ht(xi;l))
Zt
8i = 1::m; 8l = 1::k
Zt is a normalisation factor
(so that Dt+1 will be a distribution)
end for

return the combined hypothesis:
f(x; l) =XT t=1 ht(x; l)
end AdaBoost.MH

VII. Evaluation Results

VIII. Conclusion

In Online Social Networks, an intermediate system is implemented to filter undesired messages posted in private user wall. The comment posted in Online Social Networks is short text and non-English word. To classify a short text message a distributional term representation and to classify a non-English word latent semantic analysis method are used. The Filtered wall system exploits a Machine Learning Approach to enforce a customizable content dependent Filtering Rules. The flexibility of the system is in terms of filtering options which is enhanced through Block Lists. The Filtering rules and Block list is dependent on the profile details and comments posted on the user walls.

References

[1] *Adomavicius.A.,et.al., "Toward the Next Generation of Recommender Systems: A Survey of theState-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.*

[2] *Bobicev.V, Marina Sokolova, "An Effective and Robust Method for Short Text Classification," Proc. 23rd Nat'l Conf. Artificial Intelligence (AAAI), D. Fox and C.P. Gomes, eds., pp. 1444-1445, 2008.*

[3] *Chau.M.,et.al., "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] *David Meyer., Support Vector Machines ,The Interface to libsvm in package e1071,January 10, 2014*

[5] *Guillaume Lemaitre, Miroslav Radojevi , "Directed Reading: Boosting algorithms" Mach. Learn., 65(2-3):473/484, 2006.*

[6] *Juan Manuel Cabrera.,et.al., "Distributional Term Representations for Short-Text Categorization", pages 1225–1226,2012.*

[7] *Joachims.T.,et.al., "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," Proc. Int'l Conf. Machine Learning, pp. 143-151, 2009.*

[8] *Joachims.T.,et.al., "Text Categorization with Support Vector Machines:Learning with Many Relevant Features," Proc. European Conf.Machine Learning, pp. 137-142, 2010.*

[9] *Kim S. B., Rim H. C., Yook D. S. and Lim H. S., "Effective Methods for Improving Naïve Bayes Text Classifiers", LNAI 2417, 2002, pp. 414-423*

[10] *Klopotek M. and Woch M., "Very Large Bayesian Networks in Text Classification", ICCS 2003, LNCS 2657, 2003, pp. 397-406*

[11] *Lakshmana Pandian S.,et.al., " Machine Translation from English to Tamil using Hybrid Technique", International Journal of Computer Applications (0975 – 8887) Volume 46– No.16, May 2012.*

[12] *Mooney.R.J.,et.al., "Content-Based Book Recommending Using Learning for Text Categorization," Proc. Fifth ACM Conf. Digital Libraries, pp. 195-204, 2000*

[13] *Mengen Chen, Dou Shen "Short Text Classification Improved by Learning Multi-Granularity Topics" Journal of Mach. Learn. Res., 3:993–1022, 2007.*

[14] *Matt Healy.,et.al., "An assessment of case base reasoning for short text message classification" Proc. Int'l Conf. Machine Learning, pp. 143-151,2010.*

[15] *Quan Yuan, Gao Cong "Enhancing Naive Bayes with Various Smoothing Methods for Short Text Classification" ACM 978-1-4503-1230-1/12/04.*

[16] *Sarah Zelikovitz, "Transductive LSI for Short Text Classification Problems" ACM Transactionson Information Systems 12(3):252–295.*

[17] *Schneider, K., Techniques for Improving the Performance of Naive Bayes for Text Classification, LNCS, Vol. 3406, 2005, 682- 693.*

[18] *Shanahan J. and Roma N., Improving SVM Text Classification Performance through Threshold Adjustment, LNAI 2837, 2003, 361-372.*