A System to Filter Unwanted Messages from OSN User Walls AlongWith Opinion Mining

Ms. Akanksha Darvekar; Ms.Shraddha Dudhe & Ms.Darshana Kedar

Abstract :

In the recent years, Online Social Network(OSN) has become one way for the entertainment specially for the youngsters. Most of the people use it and its graph is growing day-by-day. The OSN help people to stay connected with each other, share information, views and experience with others. But they doesn't provide us to have control on our messages that will be post from us and posting of such messages annoys other. A system called Machine Learning is introduced which help to filter the undesirable messages. This system can't work with images as number of different images can be form. The aim of the present work is to filter the undesirable messages with the help of automated system called Filter Wall. Also the views are shared among the people, so , with the help of opinion mining impact of the category or subject on the people can be easily known. With the help of the datasets that are already that are fed to the system Positive , Negative and Neutral views can be computed.

Keywords-Online Social Network; Machine Learning; Filter Wall; Opinion Mining.

Introduction :

An interactive medium to share , communicate considerable amount of human life resources is OSN. Several type of text , audio ,video and other type of data is shared through it.Online Social Network is the platform to build social networks and relations among people.In each social network service each user have its own profile i.e a personal space,social links and many additional services.As it is the web based service each user can create their own profile and can also create the list of persons to which user want to connect.

But , now-a-days , it is observed that people make abusive comments on the post. They have no control over their comments which they are going to post .In order to avoid this, information filtering can be used .The private area of the user on which posting and commenting is done known as Walls. Because of information filtering, user will have control over the messages and filtering of unwanted messages would be carried out.

Today's OSN provide us with the very little support to prevent unwanted messages on user walls. For example, Facebook , provides to view our profiles only by the friends , friends of friends etc and only they can post messages on user wall. Filtered Wall is use to filter unwanted messages from OSN user walls .Machine Learning text categorization technique is used here to categorize the word according to the dataset of word that is fed to the system.These datasets are the Blacklists that is temporarily used to prevent posting of unwanted messages on user walls.

It is also observed that , people post images , videos or audios which comes or belong to the particular category. Others make comments or post opinions on it. Because of these opinions reviews of people come to know easily but it is difficult to know whether they are showing Positive , Negative or Neutral impact . So, here the concept of Opinion mining is introduced ,which shows the result in the form of pie chart and one can easily predict the result from it.
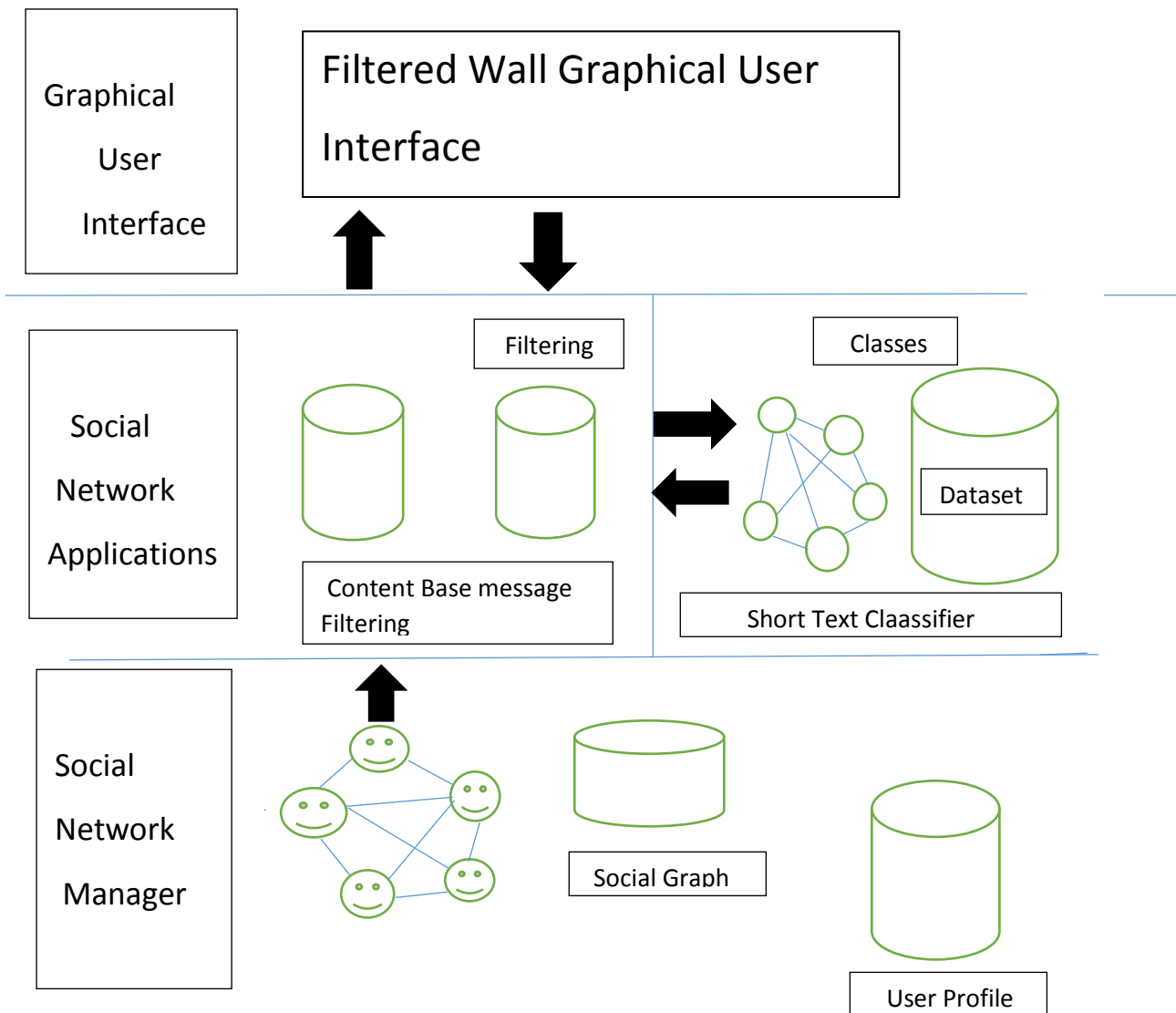
Fig 1 : OSN General Architecture

Opinion Mining:

Opinion mining refers to computational techniques for analyzing the opinions that are extracted from various sources. Current opinion research focuses on business and e-commerce such as product reviews and movie ratings.Opinion target, opinion holder and opinion are the definitions used to extracting opinions from different online sources.

An opinion can be expressed in two types. 1. Direct opinion, 2.Comparative opinion. All the opinions are stored in a document.

Following are the steps to extracting the opinions.

- Identify the objects.
- Feature extraction and synonym grouping.
- Opinion orientation determination.
- Integration

Advent of Web and social media content has much excitement and createdabundant opportunities for understandingthe opinions of the general public and constumerstoward social events, political movements, company

strategies, marketing campaigns, and product preferences.Many new and exciting social, geo political,and business-related research questions can be answered by analyzing the thousands, even millions, of comments and responses expressed in various blogs , forums (such as Yahoo Forums), social media and social network sites and tweets (Twitter).*Opinion mining*, a sub discipline within data mining and computational linguistics,refers to the computational techniques for extracting, classifying,understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated content.Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and other emotional states in online text.

EXPECTED OUTCOME

✓ Popularity Graphs
✓ Positive and negative and neutral opining mining
✓ Product review summary

The topics covered include how to extract opinion, sentiment, affect, and subjectivity expressed in text. Researchers have been able to classify text segments based on sentiment, affect, and subjectivity by analyzing positive or negative sentiment expressed in sentences, the degree of violence expressed in forum messages and so on.There are stillpromising new directions for developing and new opinion mining research. For example, much past and current opinion mining research has focused on English, Chinese, Arabic and several European languages. Advanced techniques have been developed especially for English.The large amount of public opinions expressed by citizens in different parts of the world, new, scalable opinion mining and sentiment analysis resources and techniquesneed to be developed for various languages. Future work in multilingual opinion mining willrequire bootstrapping techniques for analysing obscure and lesser-known languages for quick situation assessment. Frameworks and methods for integrating sentiments and opinions expressed with othercomputational representations like product features extracted from user-generated text, participant reply networks, spikes and outbreaks of ideas or events are also critically needed.Much of the current opinion mining research has focused on business and e-commerce applications, such as product reviews and movie ratings.
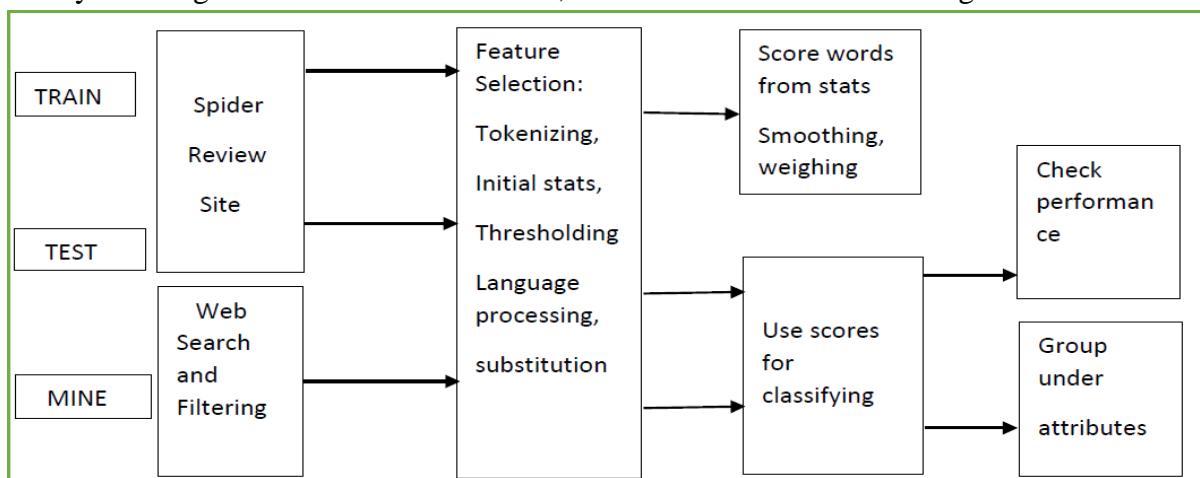

Fig 2 : General view of Opinion Mining

Existing System :

Indeed, Today OSNs provide very little support toprevent unwanted messages on user walls. For example, Face book allows users to state who is allowed to insertmessages 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 messuch as political or vulgar ones, no matter who posts them.

1.Online network facilitates its users to insert message on wall ,but no content preferences such as abusing or offensive can be suggested and thus cannot be prevented.

2.Many social networking sites support blacklisting of users. For example, blocking a user on Facebook , one can specify such user by name but no system alert on the basis of frequent bad messages seen to wall owner.

3. An account unblock cannot be done even after user prior requesting which will motivate user to go for new account creation , results into the proxy accounts.

4. Even though the social networks today, have the restrictions on the user who can post and comment and on any user wall , they do not have any restriction on what they post.so some people will use the indecent and vulgar words on commenting the user post.

Nature of Work :

Machine learning (ML) is used as text categorization techniques to automatically assign each short text message with in a set of categories based on its content. The major efforts in building a robust Short Text Classifier (STC) concentrate in the extraction and selection of a set characterizing and discriminating features. Here, a database of the categorized words is built and it is used to check the words if it has any indecent words. If the message consists of any vulgar words, then they will be sent to the Blacklists to filter out those words from the message. Finally, the message without the indecent words will be posted in the user's wall on the result of thecontent-based-filtering technique.

Advantages of Proposed System:

1. System automatically filters unwanted messages using the blacklist on the basis of bothmessage content and the message creator relationship and characteristics.

2. Major difference include , a different semantics for filtering rules to better fit the considered domain to help the users filtering rule specification , the extension of the set of features considered in the classification process.

3. As studies show that cyber bullying can occur due to undesired post of text and images ,so The motive of proposed system is to build an automated system to filter unwanted messages from OSN user wall. Enable users to have a control on the messages posted on their walls. Images will also be filtered but by user only.

4.Impact of a particular post can also be known by the reviews whose results will be analyzed by Opinion Mining.

5.Before blocking any one account at least two or three warning messages will be given and if the account is block then message will be send to user's mail.

METHODOLOGY :

Naive-Bayes Classification Algorithm

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way

by determining probabilities of the outcomes. It can solve diagnostic and predictive problems.

This Classification is named after Thomas Bayes (1702-1761), who proposed the Bayes

Theorem. Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit

probabilities for hypothesis and it is robust to noise in input data.

Uses of Naive Bayes classification:

1. Naive Bayes text classification
2. Spam filtering
3. Hybrid Recommender System Using Naive Bayes Classifier and Collaborative   Filtering
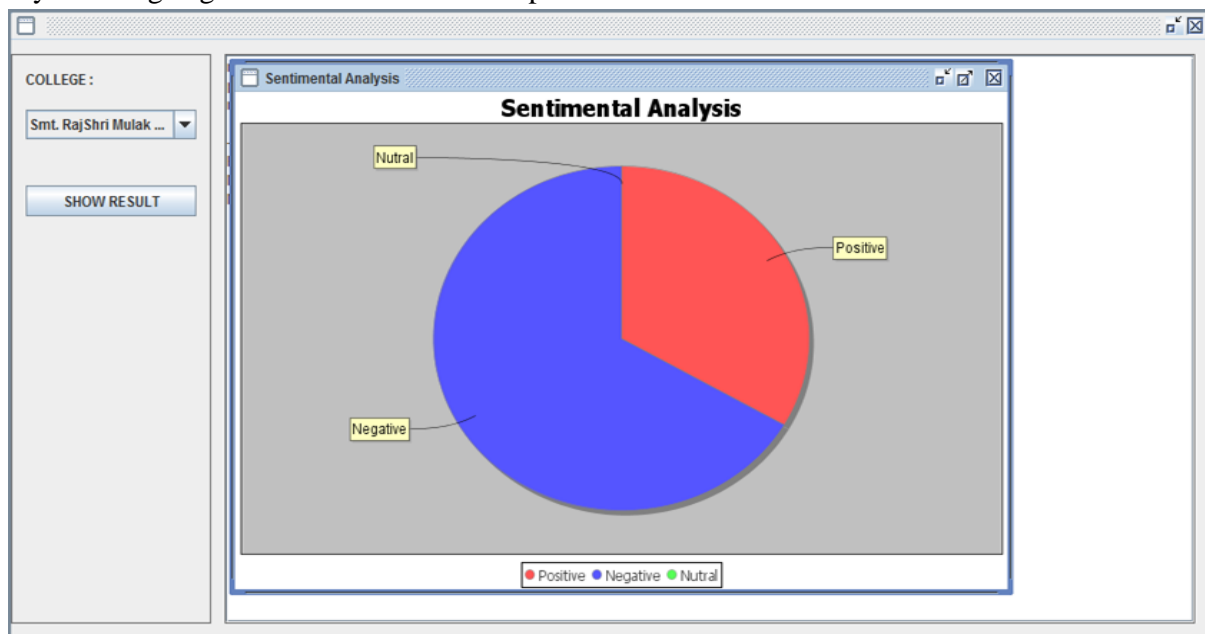4. Online applications

Fig: Sentimental Analysis

CONCLUSION :

Through our project we will be gathering reviews or opinions from people.The reviews will be as data for us on which Naïve bayesian algorithm get apply and opinion mining is done along with text categorization come into existence as a Filteration of unwanted messages in future.

REFERENCES :

[1] M. Carullo, E. Binaghi, and I. Gallo, "An online document clustering technique for short web contents," Pattern Recognition Letters, vol. 30, pp. 870–876, July 2009.

[2] M. Carullo, E. Binaghi, I. Gallo, and N. Lamberti, "Clustering of short commercial documents for the web," in Proceedings of 19ᵗʰ International Conference on Pattern Recognition (ICPR 2008), 2008.

[3] https://www.facebook.com

[4] K. Dave, S. Lawrence, and D. M. Pen opinion extraction and semantic clas *Proceedings of the 12th international* WWW '03, pages 519–528, New York

[5] R.F. Xu, , K.F. Wong, and Y.Q. Xia WIA in NTCIR-7 MOAT Task, *Proc. o*

[6]G. Wang, K. Araki,. Modifying SO-P Mining by Using a Balancing Factor an In *Proceedings of NAACL HLT 2007,* 189 – 192, Rochester, NY, 2007

[7]wamptutorials.blogspot.in,https://www.mysql.com