

Handling Different Types of Messages in On-Line Social Network using Naïve Bayes

Libina Rose Sebastian

*Department of Computer Science and Engineering
St. Joseph's College of Engineering and Technology Palai,
Kerala, India*

Smitha Jacob

*Department of Computer Science and Engineering
St. Joseph's College of Engineering and Technology Palai,
Kerala, India*

Teena Thomas

*Department of Computer Science and Engineering
St. Joseph's College of Engineering and Technology Palai, Kerala, India*

Abstract

Online Social Network is a platform to make social relationship between people who share information. Twitter, Facebook etc. are the most popular OSN we are familiar with. Various types of data can be shared such as pictures, music, video etc., through the social networking sites. We can post messages like greetings, wishes or any other type of messages on other user's wall. Spam messages are irrelevant or unsolicited messages sent over the Internet. Today's OSN suffers from the problem that unwanted messages can be posted. This should prevent to keep a user's account secure. There are many mechanisms related to this are provided but no content based mechanisms are there. In Facebook, we can block another user. The message may contain unwanted contents. If the user is not able to access the account, the message will be posted to the wall. Here a message filtering mechanism is used to keep unwanted message away from the user wall. So that the message will be posted based on user preferences.

Keywords: Data mining, Text classification, Naïve Bayes, Social Network

I. INTRODUCTION

Information filtering mechanisms are used to deal with filtering of unwanted data from a large amount using some methods. Privacy is a problem in today's social network [7]. The concept of message filtering [10] [13] and unwanted message filtering are already presented [1]. Filtered Wall (FW) is used to filter unwanted messages. It is a customized mechanism to filter messages. Machine Learning based classifiers [14] [16] [8] can be used to classify the messages into set of categories. There are mechanisms to identify the sentiment. Here Naïve Bayes classifier is used, which is a popular classifier and perform well on large data set. Text classification classifies the text into two or more categories. It is the process of automatically sorting a set of documents into categories from a predefined set. Classification can be done from short text to large documents. Text classification is commonly used to handle spam emails [3] [11] [15], classify large text collections into topical categories, and manage knowledge and also to help Internet search engines. The data set can be categorized according to the topic. There are mechanisms which is done by labeling representative words for different classes [2].

In [5], filtering is done by offensive words using grammatical retains Filtering Rules (FR) is another feature of this system, which is used to specify what contents, should be displayed and not displayed. Some users may intentionally try to post unwanted messages to another user's wall. In addition with the filtering of unwanted messages, users who post unwanted messages should also be blocked. These profiles can be marked under Blacklist (BL). When a user is under blacklist, the messages from that particular user will not be presented. This makes the system more powerful. When a user tries to post a message on the wall, it will be passed through Filtered Wall. The classifier used i.e., Naïve Bayes classifies the message into either spam or good. The message will be passed to the classifier and the system learns from the provided dataset. Machine Learning based classifiers automatically classify the text. Therefore it makes a user to not bother about other users or their messages. Here a user can post English and Malayalam text messages and audio messages.

Whenever new messages are arrived it will automatically be classified based on the trained information. This is the advantage of using the Machine Learning based classifiers. Documents with a probability greater than a certain threshold are considered relevant for this category. If the message contains unwanted contents, it will be discarded from the wall, otherwise it will be displayed. The user who posts the message can be marked as Blacklist. It is possible to mention the time period for Blacklist. If the user tries to post again after the time expires, he can be blocked permanently from posting any kind of messages in future. The system is more powerful as the system is already trained. So that whenever new messages are arrived, it identifies the class from the previously learned training set.

If the training set is large, effective classification can be provided. If a word occurred multiple times in a class, the classifier assumes the word belongs to that particular class. So that whenever new messages are entered, based on the occurrence of the word in the pre-defined class the classifier can identify membership. If the classifier needs to be trained accurately, more training

set is needed. If a word occurred in several classes, based on the number of occurrence in a class is used to identify membership of the input message.

Spam is not only a problem with online social network, SMS also contains spam [9]. More number of users go for reviews when they want to buy a product or want to watch a movie. Spam also present in reviews also. Studies are there related to this problems also [6]. Users usually use short text for communication on social networks. This is a challenge for text classification as a words may be meaningless. Short text based studies are useful to understand their features [12]. There are many other privacy related methods presented previously. All these studies shows the importance of providing secure connections in many areas.

II. DESCRIPTION

A. Message Identification

In Online Social Network users are connected with each other. A user can send messages to other users in his friend circle. Information filtering mechanisms are used to deals with filtering of unwanted data. Here the process of filtering unwanted message is explained. The system uses Naïve Bayes classifier to do the classification. Naïve Bayes classifiers can provide better results. Its simplicity increases its use in many areas. It provides good results based on the training set provided. The training is done to learn the classifier. Training samples are nothing but set of pre-classified data set. Whenever a new input is provided, based on the training provided, the classifier identifies the belonged class of that entry.

Naive Bayes classifier is based on Bayes theorem. By calculating the probability, the corresponding class of an entity is identified. This is the advantage of using classifier. The computation time to find the class is very less for Naïve Bayes classifier. Once the classifier shows the message as a spam one, it will be discarded. If the message is genuine one, it will be displayed. There is Blacklist to add profiles of users who post unwanted messages. If a user tries to post unwanted message, it will be automatically discarded and that user will be under blacklist. The messages from these users will not be displayed on the wall. Classifiers reduce the job of manual classification as it automatically classifies the text. Machine Learning based classifiers automatically learn from previous data set. It will be helpful for new set of data set. The text document can be used for different purposes like E-mail filtering, E-Commerce, Question answering etc. In email filtering, emails can be filtered based on the needs of users. Some emails contain spam messages. These kinds of messages should not be allowed to reach in users account. In ecommerce applications, set of categories can be provided for each product. This makes the easy browsing of products. When people are not aware of what they want, they usually go for browsing. So that classification is an essential one for ecommerce applications.

When a user query about an event or topic, answers related to that particular question should be given. For example when someone ask “how to delete Facebook account?” answers should be related to deleting Facebook account. If the answers are not related to the question, then the users will not be satisfied with that. These features can be provided only based on classification. Topic based text classification uses topics to classify the text. Many text classifiers use Machine Learning techniques. Naïve Bayes is used widely because of its simplicity and effectiveness. It is used in large documents and provide good results and high speed.

B. Classification

The input text is classified into different categories. The classifier uses set of text files and get trained with that. When new messages are entered, it automatically classify the texts based on the training. This is the advantage of using classifiers. Based on the training set we are provided, the classifier get learned. The strength of each word in the text documents are important. When a word appear multiple times in a text document, then the classifier understands that the word belongs to a particular class. Set of pre-classified data set are provided to classify the text documents. The text is classified based on these training set. When the user enters the message in the text area provided, these messages are compared with the training set. This is done automatically as the classifier learns from the training set. This is the advantage of using the training set. Here Naïve Bayes classifier is used to classify the input text. Which is a popular classifier and provide good result. It is widely used in many applications because of its simplicity and effectiveness. Naïve Bayes classifier classify the input text into a particular class “genuine” or “spam”. In order to understand the membership of the text in a class manual classification is needed. This manually classified data set are provided as training set.

The user can enter the message in the space provided. A text area is provided to enter the input message. There are different types of messages like good and spam. This input text is used to compare with the text files provided. Text classifiers are provided to classify the given input text. A user can provide any kind of message to the space provided. It may be a genuine message or a spam message. Input text is an important factor for text classification. The provided input text will be compared with the training set.

C. Training

The classification algorithm builds the classifier by analyzing or “learning from” a training set. Set of pre-classified data set are provided in order to learn. When a new input given, the classifier can identify the class based on the training. The presence of a word in the training set are helpful for finding the membership of text. If a message is presented in more than one classes, the

IV. CONCLUSION

Online social network allows users to share information. The posting of unwanted messages from users are a major problem in today's online social networks. Prevention of such message is challenge. Filtered wall is a mechanism to filter unwanted messages. It helps to make a user's wall away from such messages. After a user enters into the system he can post messages to the wall. Filtering need to be done to secure user profiles. Filtered wall allows customized filtering process. The user can specify which type of message should be banned from his wall. The existing systems not provided these kind of content based filtering. There are mechanisms to categorize messages. It categorizes different messages based on user preferences. Then a user gets the specified message only. The problem with this system is that no filtering is there. Email can be filtered as it contains unwanted spam messages. Based on the type of message, it will be categorized as spam and legitimate emails. The spam message will not be present in the inbox.

Spam messages are present in different form. There are many review sites present and spam messages can be posted. Those kinds of messages also should be removed in order to make secure browsing. Mechanisms to resolve these kinds of messages are presented. Providing secure browsing is another challenge. Content based filtering can remove messages of spam type. The Machine Learning based filters can identify the messages based on contents. Preventing users to post any kind of spam messages can be done based on available techniques.

Naïve Bayes classifier supported Malayalam fonts also. Different languages can be included so that a user who is familiar with his mother tongue can be easily sent messages. The concept of filtering unwanted messages can be applied to images also. As today's social networks uses different types of message like audio, video, image etc. the type of message is also important. Here audio is also considered. Audio message from the sender to receiver will be converted as texts and then the classification is performed. Then the text will be passed as audio to receiver. The surrounding noises affect the input.

REFERENCES

- [1] Vanetti, Marco, et al. "A System to Filter Unwanted Messages from OSN User Walls." *Knowledge and Data Engineering, IEEE Transactions on* 25.2 (2013): 285-297.
- [2] Liu, Bing, et al. "Text classification by labeling words." *AAAI*. Vol. 4. 2004.
- [3] Youn, Seongwook, and Dennis McLeod. "A comparative study for email classification." *Advances and innovations in systems, computing sciences and software engineering*. Springer Netherlands, 2007. 387-391.
- [4] Bhargavi, P., and S. Jyothi. "Applying Naive Bayes data mining technique for classification of agricultural land soils." *International journal of computer science and network security* 9.8 (2009): 117-122.
- [5] Xu, Zhi, and Sencun Zhu. "Filtering offensive language in online communities using grammatical relations." *Proceedings of the Seventh Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*. 2010.
- [6] Jindal, Nitin, and Bing Liu. "Analyzing and detecting review spam." *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*. IEEE, 2007.
- [7] Beye, Michael, et al. "Literature overview-privacy in online social networks." (2010).
- [8] Baharudin, Baharum, Lam Hong Lee, and Khairullah Khan. "A review of machine learning algorithms for text-documents classification." *Journal of advances in information technology* 1.1 (2010): 4-20.
- [9] Delany, Sarah Jane, Mark Buckley, and Derek Greene. "SMS spam filtering: methods and data." *Expert Systems with Applications* 39.10 (2012): 9899-9908.
- [10] Belkin, Nicholas J., and W. Bruce Croft. "Information filtering and information retrieval: two sides of the same coin?" *Communications of the ACM* 35.12 (1992): 29-38.
- [11] Christina, V., S. Karpagavalli, and G. Suganya. "A Study on Email Spam Filtering Techniques." *International Journal of Computer Applications* 12.1 (2010): 0975-8887.
- [12] Sriram, Bharath, et al. "Short text classification in twitter to improve information filtering." *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2010.
- [13] Gao, Hongyu, et al. "Towards Online Spam Filtering in Social Networks." *NDSS*. 2012.
- [14] Blanzieri, Enrico, and Anton Bryl. "A survey of learning-based techniques of email spam filtering." *Artificial Intelligence Review* 29.1 (2008): 63-92.
- [15] Shrivastava, Jitendra Nath, and Maringanti Hima Bindu. "E-mail spam filtering using adaptive genetic algorithm." *International Journal of Intelligent Systems and Applications (IJISA)* 6.2 (2014): 54.
- [16] Ikonomakis, M., S. Kotsiantis, and V. Tampakas. "Text classification using machine learning techniques." *WSEAS Transactions on Computers* 4.8 (2005): 966-974.