Volume 3, Issue 11, November 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Evaluation of Social Networking Site Conversation Filtering Based on Bayesian Theory

Shreya Mahida*, Arpit Trivedi
Smt. Chandaben Mohanbhai Patel
Institute of Computer Applications, CHARUSAT, India

Abstract---In this paper, the system that classified regular chat and improper chat (offensive conversation) was constructed by filter with Bayesian theory used well by the text classification task as a text filtering algorithm. It was confirmed to evaluate the performance of the text filter constructed by Bayesian theory and to show a relevance ratio. Moreover, the text filtering method was built into Bayesian text filter and the relevance ratio was able to be improved for social networking sites chat or conversations.

Key words: Bayesian theorem, Text filtering, Text classification, Relevance Ratio, Probability

I. INTRODUCTION

Millions of people use social networking sites every day, but main problem in social networking sites is offensive chat in which communication message contain harmful words or appalling words. Text filters are necessary to control on sending and receiving such kind of offensive words, because the content of chat conversation is basically described by the text. It can be said that task of classifying text into regular chat and offensive chat is text classification task. Therefore, various text classification algorithms can be applied for the text classification task. The goal of text classification is to classify set of sentences into predefined category. Example: to classify sentences either offensive or regular sentence. Bayesian algorithm is easy to implement for probability based filtering and classifying a sentence. Text filtering is used to remove offensive words from the sentence and generate relevance ratio and reproduction ratio of the words. In this paper, the system that classified offensive chat and regular chat was constructed by filter with Bayesian theory used well by the text classification task as a text classification algorithm. Bayesian approach is the statistical based text filter method which is strong algorithm for classification in which probability is calculated. Particular word has particular probability of occurring text. The probability of a certain cause when a certain event occurs can be calculated by the probability of all cause of event.

II. CONVERSATION CLASSIFICATION AND FILTERING TASK

In social networking site the conversation can be done via many forms (text, images, music etc). In this paper we are going to focus on the text based conversation for social networking sites and classify the text in terms of formal(regular) and informal(offensive) words.

Our goal is to have our filter with

1. List every word from chat conversation
2. Determine words appearing in chat conversation
3. Use those words as an input for Bayesian formula to determine if the word is regular or offensive
4. Generate the relevance ratio

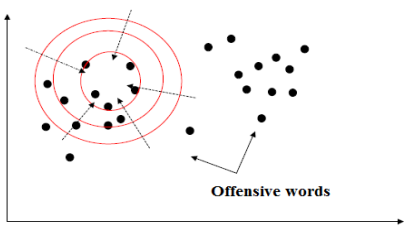

Fig. 1: Text filtering

III. BAYESIAN TEXTFILTER

Bayesian text filter is a conversation filter which is based on Bayesian Theory. In Bayesian theory the probability of a certain cause when a certain event occurs can be calculated by the probability of all cause of event and conditional probability that event occurs by certain cause. The filter separates by the probability of whether the offensive

chat or regular chat from the appearance probability of word (token) used with chat conversation based on Bayesian theory.

When token (*w*) is included the probability that the conversation is offensive (offensive sentence probability: p (*w*)) is defined by following expression

$$P(w) = \frac{P(b)/n_{bad}}{a \,.P(g)/ n_{reg} + P(b)/n_{bad}}$$

In this expression the symbols are defined as follows

- P(*w*) : When a token(*w*) is included, that the probability of sentence is offensive sentence(offensive sentence probability)
- n_{bad}: number of bad words
- n_{reg}: number of regular words
- P(*b*): probability of bad words
- P(*g*): probability of regular words
- a : total number of words in sentence (weight)

In this definition the mis-detection rate of offensive words is decreased by applying weight to total number of regular words.

The Bayesian filter technique for filtering sentence weather the sentence is offensive or regular is as follows. The technique divides into number of stages

- Preprocessing(Filter Learning)
- Classifying(Filtering Process)

Preprocessing (Filter learning)

1. Assemble chat conversations
2. Divide into tokens
3. Calculate offensive sentence probability for each token

Classifying (Filtering)

1. Divide conversation in to token
2. Query offensive sentence probability(OSP) of token
3. If OSP greater than zero than sentence is offensive sentence
4. If OSP is equal to zero than sentence is regular sentence
5. Based on the OSP the relevance ratio of offensive words and regular words for a chat conversation is formed

IV. IMPLEMENTATION OF TEXT FILTER

We implement the Bayesian text filter and evaluate the performance. The relevance ratio is used for performance evaluation. Relevance ratio is defined as follows;

$$rel = s/n$$

In this expression the symbols are defined as follows:

rel: relevance ratio
s: total number of bad words in sentence
n: total number of words in sentence

V. IMPLEMENTATION OF BAYESIAN TEXT FILTER

BT filter is used the alphabets (A-Z), numbers (0-9), special characters and delimiters. We took 5 offensive chat conversations were prepared and performance was evaluated by cross validation method. The experiment result is as shown in following table:

Source	**Relevance Ratio (%)**
Offensive Words	45%
Regular Words	55%

VI. CONCLUSION

In this paper, the system that classifies the chat conversation in social networking sites is either offensive or regular was constructed by Text filter with Bayesian theorem used well by text classification as a text filtering algorithm. It is confirmed to evaluate the performance of text filter constructed by Bayesian theory and to show relevance ratio. Using this method we can prevent offensive chat conversation in social networking site. It can be concluded that the performance of the text filter can be improved with help of other machine learning algorithms.

REFERENCES

[1] Y., Hasegawa, T., Watanabe, I., Sato Ichimura, "Text Mining case Study," *Journal Japanese Society for Artificial Intelligence*, vol. 16, no. 2, pp. 192-200, 2001.

[2] M., Taira Nagata, "Text Classification - Showcase of learing Theories," *IPSJ Magazine*, vol. 42, no. 1, pp. 32-37, 2001.

[3] Hirofumi Inomata, Masaki Miyamoto and Osamu Konishi Ayahiko Niimi, "Evaluation of Bysian Spam filter and SVM spam filter".

[4] Kawano , H., Arimura,H Naskawa T., "Base Technology for Text mining," *Journal Of Japanese Society for Artificical Intelligence*, vol. 16, no. 2, pp. 201-211, 2001.

[5] Durga Torshniwal Arun Rjput, "Adaptive Spam filtering based on Bayesian Algorithm," *International Journal on Advanced Computer Engineering and Communication Technology*, vol. 1, no. 1, pp. 8-11.

[6] nabeken. bsfilter/ bayesian spam filte. [Online]. http://www.h2.dion.ne.jp/ nabeken/bsfilter/

[7] E.Binaghi,B.Carminati,M.Carullo and E.Ferrari M.Vianetti, "Content Based Filtering in online socical network".

[8] Sathiyakumari.K Padma Priya.B, "Classification of Unwanted Messages in Online Social Network Using Machine Learning Algorithms," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 4, no. 8, pp. 2697-2701, Aug 2013.

[9] John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003) Andrew Gelman, *Bayesian Data Analysis*, 2nd ed.: CRC press, 2003.

