Content based Hybrid SMS Spam Filtering System

T. Chaminda[1], T.T. Dayaratne[2], H.K.N. Amarasinghe[3], J.M.R.S. Jayakody[4]
Faculty of Information Technology, University of Moratuwa, Sri Lanka
thilakc@uom.lk[1], thusithathilina@gmail.com[2], hknamarasinghe@gmail.com[3], jayakody1116@gmail.com[4]

Abstract - **World has changed. Everybody is connected. Almost each and everyone have a mobile phone. Millions of SMSs are going around the world over mobile networks in every second. But about 1/3 of them are spam. SMS spam has become a crucial problem with the increase of mobile penetration around the world. SMS spam filtering is a relatively new task which inherits many issues and solutions from email spam filtering. However it poses its own specific challenges. Server based approaches and Mobile application based approaches are accommodate content based and content less mechanism to do the SMS spam filtering. Though there are approaches, still there is a lack of a hybrid solution which can do general filtering at server level while user specific filtering can be done on mobile level. This paper presents a hybrid solution for SMS spam filtering where both feature phone users as well as smart phone users get benefited. Feature phone users can experience the general filter while smart phone users can configure and filter SMSs based on their own preferences rather than sticking in to a general filter. Server level solution consists of a neural network along with a Bayesian filter and device level filter consists of a Bayesian filter. We have evaluated the accuracy of neural network using spam huge dataset along with some randomly used personal SMSs.**

Key words – SMS, Spam, Neural Network, Bayesian Filter

I INTRODUCTION

SMS spam is a real and growing problem primarily due to the availability of very cheap bulk pre-pay SMS packages and the fact that SMS engenders higher response rates as it is a trusted and personal service. SMS spam is defined as any unwanted SMSs received on a mobile device. SMS spamming is becoming a leading issue among the millions of mobile phone users recently. Most of the spam messages are being sent by marketing companies that are trying to find people who will respond so they can sell those people's details to claims or debt management firms. Apparently these companies sending the SMSs are randomly generating mobile telephone numbers. Some companies may choose to send SMS messages rather than email, because many email spam filters will block their messages.

In recent years, spam SMSs comes from the mobile operators themselves. Most of us experienced with lots of SMSs coming from the mobile operator by informing new features ring/ringing tone promotions etc… Despite the fact that this is a kind of spam, the information in these messages still has a meaning for some users. Therefore SMS filters should have feature to identify those important texts and detect other spam SMSs that are coming from unsolicited electronic communication from third parties such as financial institutions, retailers and stores. Therefore, people's opinions about the role of the spam are divided roughly equally.

The spam SMSs appear to breach the Privacy and Electronic Communications Regulations because they are being sent to individuals without prior consent and without identifying the sender. The SMSs also appear to violate other legislation as well. In order to get rid of this problem some of the mobile phone application developers introduced few SMS spam detecting and blocking applications.

SMS spam filtering is a relatively new task which inherits many issues and solutions from email spam filtering. However it poses its own specific challenges. There are some approaches to do the SMS spam filtering, server based approaches and Mobile application based approaches. These categories can accommodate content based and content less mechanism to do the filtering.

Naive Bayer's algorithm, pattern Matching algorithms, evolutionary algorithms, Logistic Regression (LR), Dynamic Markov Compression (DMC), etc… can be used in field. Naive Bayesian algorithm is one of the most effective approaches used in these filtering techniques. Possibility to perform spam filtering at these devices level, leading to better personalization and effectiveness with the rapid increase of computational power in smart phones. 'SMS blocker' for Android, Postman SMS blocker is some existing mobile applications to detecting and blocks the spam SMSs, but no application yet to use machine learning techniques to do the SMS spam filtering.

This paper organized as follows; section 2 introduces some previous studies that talk about spam detection and filtering process. In section 3, an overview of the approach that been used. Section 4 provides details of design and implementation of the system. Evaluation strategy and experimental results that will be presented in section 5. Finally, conclusion and future work that will be shown in section 6.

II RELATED WORK

Mobile applications as well as server-side approaches accommodate either content based or Non-content based approaches. Experiment of applying different evolutionary

and non-evolutionary classifiers for spam filtering by examines the byte-level features of SMS at the access layer of mobile devices showed that the evolutionary classifiers can efficiently detect spam SMSs at the access layer of a mobile device [1]. Combination of Winnow algorithm with orthogonal sparse bigrams with refined preprocessing and tokenization techniques to achieve a higher accuracy while overcome limitation of the feature space [2]. Jose Maria Gomez Hidalgo et-al showed that the running time of learning with SVM has been comparable to Naive Bayes, and much smaller than the running time for learning rules or decision trees [3]. Structural similarity of transitory SMSs could be harnessed to cluster the short SMSs and larger clusters are indicative of bulk messaging practiced by spammers. Since algorithm clusters SMSs based on their similarity in higher dimensionality space, there is no way to determine if such vector clusters correspond to spam or legitimate SMS messages which are bulk send by algorithm itself. Given the high dimensionality short SMSs Siddharth Dixit et-al demonstrated that dimensionality reduction techniques can be utilized successfully for implementing real time SMS spam detection [4]. With the use of no of SMSs, SMS size under static features, no of SMSs during a day, etc... with graph data mining techniques Qian Xu, et-al examined the effectiveness of content-less features and showed that temporal features and network features can be effectively incorporated to build an SVM classifier, with a gain of around 8% in improvement on area under the curve compared to those that are only based on conventional static features [5].

Model of byte-level distributions of non spam and spam SMSs along with non spam and spam models using Hidden Markov Models (HMM) can be used as a robust approach to word adulteration techniques and language transformations since it works at the access layer of the mobile phone [6]. Tarek M Mahmoud and Ahmed M Mahfouz showed that use of Artificial Immune System (AIS) for filtering spam SMSs can achieve detection rate, false positive rate and overall accuracy of 82%, 6%, and 91% respectively [7].

Most of the mobile operator networks have placed anti-spoofing and faking measures which can successfully identify SMS messages that have been manipulated to forge the originating details in order to avoid charges. But the rapid increase of non spoofed or faked SMS spam messages, demand the need for a sophisticated filtering techniques. Simple server level filtering methods analyze traffic of each individual subscriber to identify high volumes of SMSs. Since spammers are using .low volumes and advanced methods, these types of simple techniques are not that much sophisticated. Wu, Wu, and Chen have used a Bayes learner to extract keywords for monitoring traffic centrally, allowing a spamminess score to be assigned, however this work was not evaluated [8]. Jie, Bei, and Wenjing have added a cost function to a Naive Bayes filter which assigned a high cost to false positives. It translates into a high spam classification threshold, and a higher threshold results in higher spam precision [9]. k-nearest neighbour algorithm (k-NN) as part of a multi-filtering approach proposed by Longzhen, An, and Longjun. After black- and white-listing, a SMS is first classified by a filter using rough sets, which provide approximate descriptions of concepts. If this filter classifies the SMS as spam, it is then passed to the k-NN classifier for final classification. An evaluation on a data set of 550 spam SMS and 200 non-spam SMS with k = 12 showed that this dual filtering method is faster and more accurate than using k-NN alone [9].

Table 1 compares the SMS spam filtering capabilities of existing android applications.

Table 1 Comparison of Android Applications

Application	Contact based	Black list	Rule based	Key word based	Machine learning
AVG	Yes	Yes	Yes	No	No
SmsBlocker	Yes	Yes	No	No	No
Quickheal	Yes	Yes	No	No	No
AntiSpamSMS	Yes	Yes	Yes	Yes	No
Numbercop	No	Yes	No	No	No
Private Box	Yes	Yes	No	N0	No
SMS Filter	Yes	Yes	No	No	No
Postman	Yes	Yes	No	Yes	No
SMS Spam Blocker	Yes	Yes	Yes	No	No
SpamBlocker	Yes	Yes	Yes	Yes	No

III Approach

Users

Each and every mobile subscriber can be considered as a potential user of the system. Further users can be categorized as smart phone users and feature phone users.

- Features phone users – Benefited by general filtering
- Smart phone users – Benefited by general filtering along with user specific filtering

Inputs

System takes SMSs from the Short Message Service Center (SMSC) via Short Message Peer to Peer (SMPP) protocol on the server level. Raw SMSs are considered as the inputs in the server level while partially filtered SMS messages are being considered as input at the device level.

Users specify categories (Jokes, News, Marketing, etc...) which are used in-order to do further filtering based on user preferences is the other input to the system.

Process

Server Level Processing

System run as a proxy between SMSC and Mobile Switching Center (MSC) and get all the SMSs that are going through the SMSC via SMPP protocol. At the server it extracts the SMS body and preprocesses the text for further processing. Spam filtering algorithms are applied on preprocessed SMSs to do the classification. If a SMS is classified as a spam SMS it will be forwarded to an email account and sending "Cancel_SM" to SMSC via SMPP in-order to cancel the normal delivery of SMS to the recipient(s). SMS messages that are classified as spam are sends to a data warehouse for future data mining purposes. If the SMS is classified as a ham without any further processing it will be forward back to the SMSC, so SMSC will handle normal SMS delivery.

Mobile Application Level Processing

SMS receiving intent in get fired as SMS received to the mobile application. System reads Protocol data units (PDU) and creates SMS messages using PDUs. It first does the filtering based on black listed numbers. If the SMS is received from a black listed number, it will be direct to the spam folder in the app and abort the broadcast of the SMSs. Bayesian filter is applies on unfiltered SMSs to do the classification. In the first runs system will not have enough data to do the classification. But it will learn from the incoming SMSs and improve with time. If the SMS gets classified as a ham it will be broadcast, so the default SMS app will handle the rest, unless it will be forwarded to the spam folder while aborting the broadcast.

Since there is a probability of ham SMSs being classified as spam, users can mark misclassified spam SMSs as not spam at the mobile application level. It helps filter to improve its' accuracy.

Output

Partially filtered SMS messages are the output at the server level while mobile app outputs further filtered SMSs based on user preferences.

IV DESIGN AND IMPLEMENTATION

Fig. 1 illustrates the top level design of the hybrid system. The most appropriate place to implement a server level spam filter is SMSC. But since SMSC's technology differs from one mobile operator to another implementing the solution as a SMSC proxy is the better solution. SMSC forward all the SMSs through the proxy application and system is capable of handling delivery of SMSs via SMPP. Fig.2 shows the top level architecture of mobile application

All the SMSs are processed in order to do the classification. System applies Bayesian filter and neural network independently to the SMSs. 10 features have been identified in the feature section phase. Features consists of character based features along with words based features. Following is the list of features that are used.

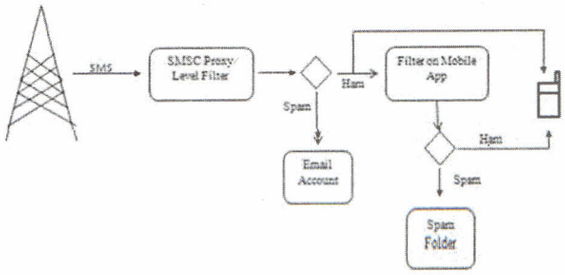

Fig. 1 Top Level Architecture of Hybrid System

1. Ratio of alpha chars to total number of characters
2. Ratio of numbers to total number of characters
3. White space ratio to total number of characters
4. Frequency of special chars (10 characters: *, _ ,+,=,%,$,@,- , \,/)
5. Frequency of punctuation 18 punctuation marks: . ‘; ? ! : ( ) – “ « » < > [ ] { }
6. Frequency of uppercase letters
7. Ration of short words to total number of words (Words having 2 or fewer characters)
8. Average word length
9. Average sentence length in characters
10. Average sentence length in words

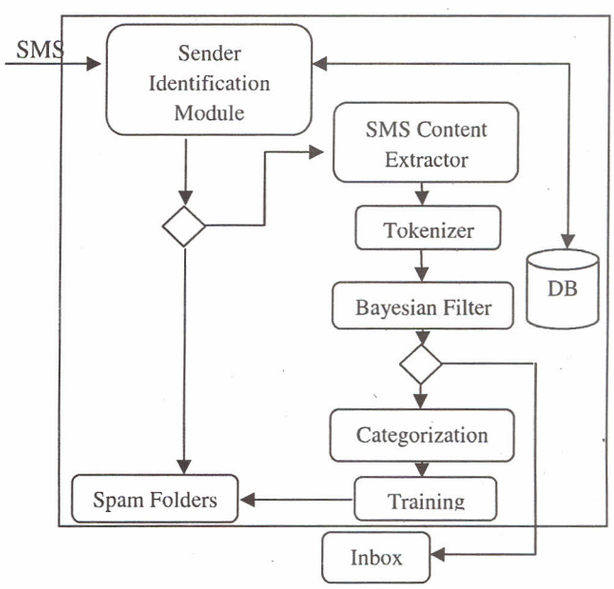

Fig. 2 Top Level Architecture of Mobile App

A neural network (NN) with 3 layers is the NN that is used in classification process at server level. Input layer consists of 10 perceptrons, where each perceptron represents a feature that were identified at the feature extraction phase. Hidden layer consists of 4 perceptrons and 1 perceptron at the output layer. Sigmoid function is

Table 2 Results comparison of data mining

No of Instances	Algorithm	Correctly Classified Instance	Incorrectly Classified Instance	Percentage of Correctly classified Instance	Percentage of Incorrectly classified Instance	Tuning
1000	J48	889	111	88.9%	11.1%	Default
1000	J48	905	95	90.5%	9.5%	Yes
1000	Part	909	91	90.9%	9.1%	Default
1000	Part	929	71	92.9%	7.1%	Yes
1000	ConjunctiveRule	848	152	84.8%	15.2%	Default
1000	ConjunctiveRule	847	153	84.7%	15.3	Yes

been used as the activation function. Results of back propagation with various learning rates and momentums as well as resilient propagation method are discussed in discussion section. System also uses Bayesian filter to do the filtering at server level. Independent outputs from the neural network as well as Bayesian filter are considered and based on that 2 results SMS is classify as either spam or ham. SMSs that are classified as spam are either send directly to user define email address or plan to save in a

First filter is trained with the SMS messages data set which includes both spam and ham SMSs. Before training the filter whole SMS is split in to words. Each and every word objects keeps its occurrences on spam SMSs and ham SMSs. When the filter trains with a spam SMS then the spam counter of its all words will be increased based on their occurrences. And also if the SMS is ham SMS then the ham counter of its all words will be increased. Based on those values, final spam probability for a word is calculated.

This application gets incoming SMSs before they go to the native inbox. Then SMSs are tested with the filter. First SMS is split into words. After that filter finds the spam probability of that words based on the stored spam probabilities. If a new word is included within the SMS then filter sets the spam value of that word to "0.5". Because there is no any detail with the filter. Then filter creates a list with the most significant words. Spam probability of each word is reduced from 0.5 and the mod value of that derived value is considered to take significant words. First fifteen Words with the highest derived values are taken to get final result.

The size of the significant words list should be kept in reasonable value. Because if words with the middle spam probabilities will direct to the wrong decisions. Here that value was taken doing some experiments and considering final spam values. After getting significant words list then filter calculates the final probabilities by applying the Bayes' rule. For that filter takes the effective spam probability (pspam) by multiplying spam probabilities of each and every word together. As well as filter calculates the effective ham probability (pham) by multiplying ham probabilities (1 – spam probability) together. Finally following equation is used to calculate final probability.

$$Final\ spam\ probability = \frac{pspam}{pspam + pham} \quad (1)$$

database where users can dial a number and get all the SMSs that are been classified as spam after sometimes.

Device level filter is a text based spam filtering application which is running on android. Training phase and the evaluation are two main phase of the application. Words are the fundamental thing that is considered within the application.

$$Final\ word\ probabilty = \frac{spam\ counter}{spam\ counter + ham\ counter} \quad (2)$$

According to above equation each and every word has its own spam probability. After the training filter will evaluate the SMS messages based on this probability.

According to the data set if this value exceeds is higher than 0.9 with spam SMSs. There for the filter marks the SMSs with the spam probability higher than 0.9 as spam SMSs.

This application has its own inbox and spam box. User can define spam SMSs according to his preference. If a SMS is wrongly classified then user can send it to the correct destination. According to that operation the filter will be updated.

User can define unwanted senders and unwanted phrases with the application. Then filter will marked SMSs as spam SMSs those are come from that given senders or included given phrases. After categorizing as spam and ham SMSs filter will further categorized those spam SMSs into several categories. They are Marketing, Service Provider, Sports and Others. Filter keeps another four filters for this categories and SMS is marked as type which gives highest spam value. This filter will train based on the user preferences and provide user specific spam filtering process.

Weka library is used for the Data mining operations. J48, Part, ConjunctiveRule are the Algorithm that are been used with WEKA. In order to do the mining and to gain high accuracy data is pre processed with StringToWordVector filter with some new configuration values.

V EVALUATION AND DISCUSSION

We have used subsets of SMS Spam Collection Data Set of 5574 SMSs [10] for training classifiers and as well as for the evaluation along with some randomly picked personal SMSs. Table 2 compares the results of various classifiers that are being used in data mining while Table 4 shows the accuracy along with the training methods that were used with the neural network. It can be seen that

Table 3 Mobile Application Evaluation

	Resulting Type																			
	Sports				Service Provider				Marketing				Other				Inbox			
	Round				Round				Round				Round				Round			
	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
Sports	2	4	5	5	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0
Service Provider	0	0	0	0	1	5	5	5	1	0	0	0	3	0	0	0	0	0	0	0
Marketing	0	0	0	0	0	0	0	0	0	5	5	5	5	0	0	0	0	0	0	0
Other	0	0	0	0	1	0	0	0	0	0	0	0	3	4	4	4	1	1	1	1

when back propagation along with 0.1 learning rate and 0.6 momentum used, it leads to the maximum accuracy level. Accuracy of the hybrid system can't be measure since mobile application accuracy is depend on users' preference, we have separately test the accuracy of mobile application. We used four types of messages as Service Provider, Sports, Marketing and Other. Table 3 results show how they were categorized with the filter. We have tested same data set in several times with the application. Mobile We haven't evaluate the accuracy of mobile application based spam filter and the effectiveness of data mining approach to the SMS spam filtering context.

Table 4 Neural Network Evaluation

Training Method	Learning Rate	Momentum	Accuracy
Back propagation	0.1	0.6	95.90%
	0.1	0.8	93.85%
	0.4	0.6	95%
	0.4	0.8	83.65%
Resilient propagation			87.50%
Manhattan propagation			76.61%

This paper presents a hybrid system for SMS spam filtering. Feature phone users can get the benefit of having a spam filter on their mobile, while smart phone user can experience the user preference based filtering using the mobile solution.

Though the smart phones getting more power, still there are limitations to do computation incentive processes at the device level.

VI CONCLUSION & FUTURE WORK

More enhanced and accurate SMS spam filtering can be obtain through the use of Neural network, Data mining, Bayesian filter. It is clear that above 90% accuracy can be obtained with the minimal changes to the existing classification techniques with the effective feature set.

Principal component analysis can be use in the future to improve the effectiveness of the features and obtain higher accuracy of the neural network. Further it is possible to do a overall evolution of the hybrid solution in order to give a better idea about the system performance.

REFERENCES

[1] - M. Zubair Rafique, Nasser Alrayes, Muhammad Khurram Khan: *Application of evolutionary algorithms in detecting SMS spam at access layer. GECCO 2011: 1787-1794*

[2] - Christian Siefkes, Fidelis Assis, Shalendra Chhabra and William S. Yerazunis, "Combining Winnow and Orthogonal Sparse Bigrams for Incremental Spam Filtering". *Proc. 15th European Conference on Machine Learning,* 2004

[3] - Jose Maria Gomez Hidalg, Guillermo Cajigas Bringas, Enrique Puertas Sánz, "Content Based SMS Spam Filtering" *Proc. ACM Symposium on Document Engineering, Amsterdam, Netherlands,* October 10-13, 2006; 01/2006.

[4] - Siddharth Dixit, Sandeep Gupta, and Chinya V. Ravishankar, "LOHIT: An On-Line Detection and Control Scheme for Cellular Spam", *Proc. IASTED International Conference on Network Security Phoenix, AZ,* Nov. 14--Nov. 16

[5] - Qian Xu, Evan Wei Xiang, Qiang Yang, Jiachun Du, Jieping Zhong, "SMS Spam Detection Using Noncontent Features," *IEEE Intelligent Systems,* vol. 27, no. 6, pp. 44-51, Nov.-Dec. 2012, doi:10.1109/MIS.2012.3

[6] - Z. Rafique and M. Farooq. "SMS spam detection byoperating on byte-level distributions using hiddenmarkov models (HMMs)" *Proc.Virus Bulletin International Conference, VB* , September 2010

[7] - Tarek M Mahmoud, Ahmed M Mahfouz. "SMS Spam Filtering Technique Based on Artificial Immune System" *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 2, No 1, March 2012

[8] - Wu, N., Wu, M., Chen, S. "Real-time monitoring and filtering system for mobile SMS" *Proc. 3rd IEEE conference on industrial electronics and applications* pp. 1319–1324, 2008

[9] - Jie, H., Bei, H., Wenjing, P." A Bayesian approach for text filter on 3G network" *Proc. 6th international conference on wireless communications networking and mobile computing,* pp. 1–5, 2010

[10] "SMS Spam Collection Data Set." [Online]. Available: http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection. [Accessed: 26-Sep-2013].