SMS Spam Filtering Technique Based on Artificial Immune System

Tarek M Mahmoud[1] and Ahmed M Mahfouz[2]

**[1] Computer Science Department, Faculty of Science, Minia University
El Minia, Egypt**

**[2] Computer Science Department, Faculty of Science, Minia University
El Minia, Egypt**

Abstract

The Short Message Service (SMS) have an important economic impact for end users and service providers. Spam is a serious universal problem that causes problems for almost all users.
Several studies have been presented, including implementations of spam filters that prevent spam from reaching their destination. Naïve Bayesian algorithm is one of the most effective approaches used in filtering techniques. The computational power of smart phones are increasing, making increasingly possible to perform spam filtering at these devices as a mobile agent application, leading to better personalization and effectiveness. The challenge of filtering SMS spam is that the short messages often consist of few words composed of abbreviations and idioms. In this paper, we propose an anti-spam technique based on Artificial Immune System (AIS) for filtering SMS spam messages. The proposed technique utilizes a set of some features that can be used as inputs to spam detection model. The idea is to classify message using trained dataset that contains Phone Numbers, Spam Words, and Detectors. Our proposed technique utilizes a double collection of bulk SMS messages Spam and Ham in the training process. We state a set of stages that help us to build dataset such as tokenizer, stop word filter, and training process. Experimental results presented in this paper are based on iPhone Operating System (iOS). The results applied to the testing messages show that the proposed system can classify the SMS spam and ham with accurate compared with Naïve Bayesian algorithm.

Keywords: *Short Message Service (SMS), Naïve Bayesian algorithm, Anti-Spam, Artificial Immune System (AIS), Tokenizer, Filter.*

1. Introduction

Short Message Service (SMS) is a popular means of mobile communication. Smart phones have become commonplace during the past few years, integrating multiple wireless networking technologies to support additional functionality and services. It was designed as part of Global System for Mobile communications (GSM), but is now available on a wide range of network standards such as the Code Division Multiple Access (CDMA) [1].

As the popularity of smart phones surged, frequent users of text messaging began to see an increase in the number of spam commercial advertisements being sent to their telephones through text messaging. Recently, we have witnessed a dramatic increment in the volume of SMS spam [2].

Spam generally refers to unsolicited and unwanted SMS, usually transmitted to a large number of recipients. SMS spam has an important economic impact to end users and service providers. The importance of increasing of this problem has motivated the development of a set of techniques to fight it [2]. The SMS spam has a bigger effect on users than email spam because users look at every SMS they receive, so SMS spam influences the users directly. Among the approaches developed to stop spam, filtering is an important and popular one. It can be defined as automatic classification of messages into spam and non-spam SMS. The challenge of filtering SMS spam is that short messages often consist of few words and sometimes these words composed of abbreviation and idioms [3].

The immune system [4] is a complex network of organs and cells responsible for the organism’ s defense against alien particles. One of the main features of the immune system is its capacity to distinguish between self and non-self genes.

In this paper, an anti-spam filtering technique based on Artificial Immune System (AIS) is proposed. The proposed technique utilizes a set of some features that can be used as inputs to a spam detection model. The idea is to classify message using trained dataset that contains Phone Numbers, Spam Words, and Detectors. Our proposed technique utilizes a double collection of bulk SMS messages Spam and Ham in the training process to improve its efficiency. To the best of our knowledge, the current work is the first SMS spam filter based on AIS classifier used for mobile devices.

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 Short Message Service (SMS) that will be shown. Section 4 provides some details of spam. In Section 5 AIS mechanism that will be describe. Section 6 contains the proposed technique. Evaluation strategy and experimental results that will be presented in section 7. Finally, conclusion and future work that will be shown in section 8.

2. Related work

Content-based filtering solutions have been proved to be effective against emails, which are typically larger in size compared to SMS messages. Abbreviations and acronyms are used more frequently in SMS messages and they increase the level of ambiguity. This makes it difficult to adopt traditional email spam filters without any modification. Healy et al. [5] discuss the problems of performing spam classification on short messages by comparing the performance of the well-known K-Nearest-Neighbor (KNN), Support Vector Machines (SVM), and Naive Bayes classifiers. They conclude that, for short messages, the SVM and Naïve Bayes classifiers substantially outperform the KNN classifier; and this contrasts with their previous results obtained for longer emails. Hidalgo et al. [6] also carried out content filtering experiments with English and Spanish spam SMS corpora to prove that Bayesian filtering methods are still effective against spam SMS messages. Gómez et al [7] proposed a content SMS spam filtering based on Bayesian filters used in stopping email spam. They analyzed to what extent Bayesian filtering techniques used to block email spam, can be applied to the problem of detecting and stopping SMS spam. Peizhou et al [8] proposed another method to filter SMS spam. They utilized Completely Automated Public Turing test to tell Computers and Human Apart (CAPTCHA) method to filter SMS spam. If the SMS can pass the CAPTCHA, it will be identified as legitimate SMS and transmitted by short message processing center. Conversely, if the SMS cannot pass the CAPTCHA, it will be identified as SMS spam and deleted by Short Message processing Center.

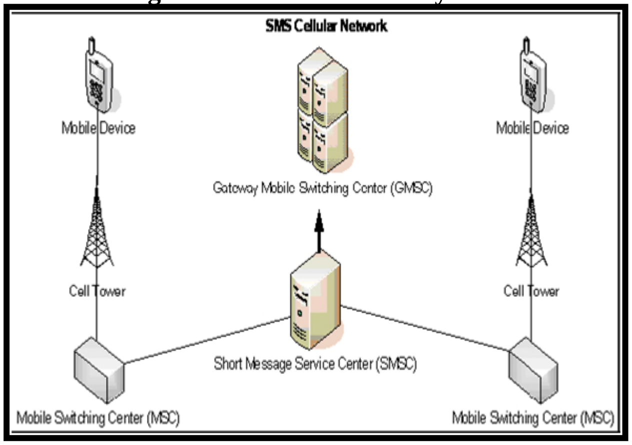
One of the drawbacks of existing solutions, however, is that they often look for topical terms or phrases such as 'free' or 'viagra' to identify spam messages. In consequence, some of the legitimate SMS messages that contain such black list words classified by mistake as spam. This could happen more frequently with SMS messages than with emails due to their smaller size and simpler content. Moreover, adaptive schemes are fundamentally weak against innovative attacks where strategies constantly evolve to manipulate classification rules. Filtering alone will not be sufficient to detect spam [9].

Many solutions against email spam have been suggested based on AIS and other techniques [3]. Most of them can effectively be transferred to the problem of SMS spam. Sarafijanovic and Le Boudec [10] proposed an AIS-based collaborative filter, which attempts to learn signatures of patterns typical of Spam messages, by randomly sampling words from a message and removing those that also occur in legitimate messages. This allows the system to be robust to obfuscation based on random words. It also carefully selects the signatures that will be distributed to other agents, to prevent the use of those relating to unreliable features. In experiments with the SpamAssassin corpus, it verified that good results can be obtained when relatively few servers collaborate, and that the proposal is robust to obfuscation.

3. Short Message Service (SMS)

SMS is a communication service standardized in the GSM mobile communication systems; it can be sent and received simultaneously with GSM voice, text and image. This is possible because whereas voice, text and image take over a dedicated radio channel for the duration of the call, short messages travel over and above the radio channel using the signaling path [1]. Using communications protocols such as Short Message Peer-to-Peer (SMPP) [11]. It allows the interchange of short text messages between mobile telephone devices as shown in Figure 1 that describes traveling of SMS between parties.

Figure 1 the basic of SMS system


SMS Cellular Network
Mobile Device
Mobile Device
Gateway Mobile Switching Center (GMSC)
Cell Tower
Cell Tower
Short Message Service Center (SMSC)
Mobile Switching Center (MSC)
Mobile Switching Center (MSC)

SMS contains some meta-data [1]:

1. Information about the senders (service center number, sender number)
2. Protocol information (protocol identifier, data coding scheme)
3. Timestamp

SMS messages do not require the mobile phone to be active and within range, as they will be held for a number of days until the phone is active and within range. SMS transmitted within the same cell or to anyone with roaming capability. The SMS is a store and forward service, and is not sent directly but delivered via an SMS Center (SMSC). SMSC is a network element in the mobile telephone network, in which SMS is stored until the destination device becomes available. Each mobile telephone network that supports SMS has one or more messaging centers to handle and manage the short messages [1].

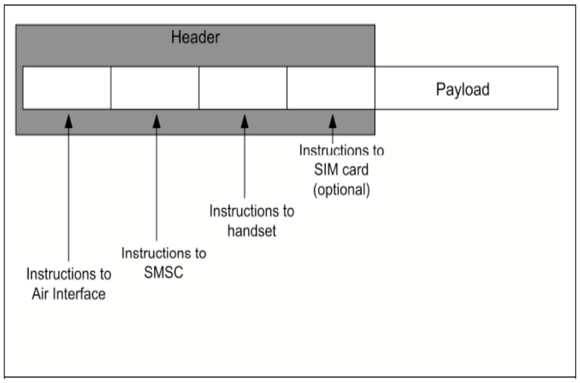

Header
Payload
Instructions to SIM card (optional)
Instructions to handset
Instructions to SMSC
Instructions to Air Interface

Figure 2 SMS message structure

As illustrated in Figure 2, the SMS comprises of the following elements, of which only the user data displayed on the recipient’s mobile device [12]:

- Header - identifies the type of message:
  1. Instruction to Air interface
  2. Instruction to SMSC
  3. Instruction to Phone
  4. Instruction to Subscriber Identity Module (SIM) card
- User Data - the message body (payload).

As shown in Table 1, each SMS is up to 140 bytes, which represents the maximum SMS size. Each short message is up to 160 characters in length when Latin alphabets are used, where each character represented by 7 bits according to the default alphabet in Protocol Data Unit (PDU) format. The length of SMS message is 70 characters in the case of using non-Latin alphabets such as Arabic and Chinese where each character represented by 16-bit Unicode format [1, 11].

Table 1: Relation between coding scheme and text length.

Coding scheme	**Text length per message segment**
8-bit data	140 byte
GSM alphabet, 7 bits	160 characters
Unicode, 16 bits	70 omplex characters

4. Spam

There exist various definitions of what spam is and how it differs from legitimate mail. The shortest among the popular definitions characterizes spam as “unsolicited bulk email”. Sometimes the word commercial added, but this extension is debatable. Another widely accepted definition states that “Internet spam is one or more unsolicited messages, sent or posted as part of a larger collection of messages, all having substantially identical content”[13, 14, 15].

Mobile spam, also known as SMS spam, is a subset of spam that involves unsolicited advertising text messages sent to mobile phones through the SMS. One of the biggest sources of SMS spam is number harvesting carried out by Internet sites offering "free" ring tone downloads. In order to facilitate the download, users must provide their phones' numbers; which in turn used to send frequent advertising messages to the phone. Wording in the sites terms of service make this legal; and users may have to go as far as to change their cell phone numbers to stop the spam.

Mobile spam problem is a much more serious problem than email spam. Mobile phones perceived as very personal devices constantly by one’s side. In addition, the costs associated per SMS are significant. As opposed to email spam where the nuisance experienced on reading it, mobile spam instantly intrudes into users’ privacy by forcefully registering its arrival. People may have several email accounts, but carry only one mobile device.

SMS spam differs from email spam in characteristic attributes. Email spam is generally identifiable by the key words used, and its structure, so that it is identifiable by various methods [16]. Table 2 illustrates some differences between email and SMS [17].

Table 2 Differences between email and SMS

Feature	Email	SMS
Length	Unlimited	160 English characters or 70 Arabic and Chinese
Process	not real-time	real-time
Representation	texts, images, attachment, etc	only texts

With the spread of SMS spam, some Mobile Network Operators have taken steps to resist spammers, and they want to reduce the volume of spam and satisfy their customers [8]. Another approach to reducing SMS spam that offered by some carriers involve creating an alias address rather than using the cell phone's number as a text message address. Only messages sent to the alias delivered; messages sent to the phone's number discarded. These solutions are not practical and does not apply on mobile agent and do not take user feedback in classification process. The computational power of mobile phones and other devices are increasing, making increasingly possible to perform spam filtering at the devices, leading to better personalization and effectiveness [9].

5. Artificial Immune System (AIS)

Artificial Immune System (AIS) is a paradigm of soft computing which motivated by the Biological Immune System (BIS). It based on the principles of the human immune system, which defends the body against harmful diseases and infections. To do this, it must perform pattern recognition tasks to distinguish molecules and cells of the body (self) from foreign ones (non-self). AIS inspire the production of new ideas that could be used to solve various problems in computer science, especially in security field. BIS based around a set of immune cells called lymphocytes comprised of B and T cells. On the surface of each lymphocyte is a receptor and the binding of this receptor by chemical interactions to patterns presented on antigens which may activate this immune cell. Subsets of the antigens are the pathogens, which are biological agents capable of harming the host (e.g. bacteria). Lymphocytes created in the bone marrow and the shape of the receptor determined by the use of gene libraries. These are libraries of genetic information, parts of which concatenated with others in a semi-random fashion to code for a receptor shape almost unique to each lymphocyte. The main role of a lymphocyte in AIS is encoding and storing a point in the solution space or shape space. The match between a receptor and an antigen may not be exact and so when a binding takes place it does so with strength called an affinity. If this affinity is high, the antigen included in the lymphocyte's recognition region [4, 10].

Clonal selection and expansion is the most accepted theory used to explain how the immune system copes with the antigens. In brief, the Clonal selection theory states that when antigens invade an organism, a subset of the immune cells capable of recognizing these antigens proliferate and differentiate into active or memory cells. The fittest clones are those, which produce antibodies that bind to antigen best (with highest affinity). The main steps of Clonal selection algorithm can be summarized as follows [18]:

Algorithm 1: Clonal selection

Step 1: For each antibody element
Step 2: Determine its affinity with the antigen presented
Step 3: Select a number of high affinity elements and reproduce (clone) them proportionally to their affinity.

6. The Proposed SMS Spam Filtering Technique

The proposed technique identifies spam on the local phone with several features to block it. These features can be described as following:

- **Black list phone numbers:** This list contains all phone numbers that the user wants to block them. In this case, the proposed technique will block the incoming SMS messages that match these numbers.
- **Black list words:** This list contains all words (spam words) that the user wants to block them. In this case, the proposed technique will block the incoming SMS messages that match these words.

Black list detectors: This list contains all detectors that built from the training process and the user feedback. The proposed system starts to analyze the incoming SMS and determine if it spam or not according to the affinity ratio between the incoming SMS and detectors list. In this case, the proposed technique will block the incoming SMS messages that match these detectors.

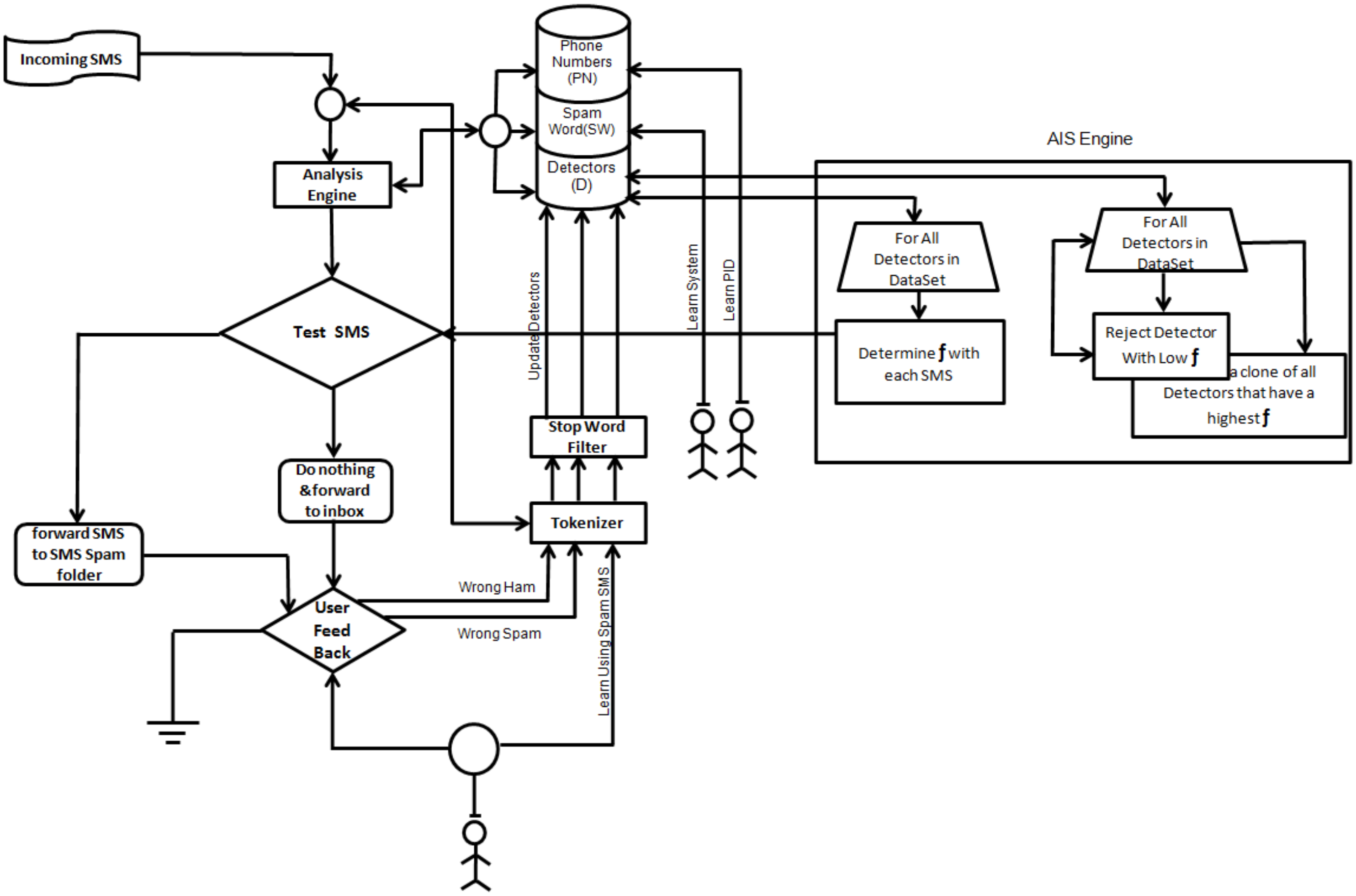

Incoming SMS
Phone Numbers (PN)
Spam Word(SW)
Detectors (D)
Analysis Engine
AIS Engine
For All Detectors in DataSet
Determine f with each SMS
For All Detectors in DataSet
Reject Detector With Low f
a clone of all Detectors that have a highest f
Test SMS
Update Detectors
Learn System
Learn PID
Stop Word Filter
Do nothing &forward to inbox
forward SMS to SMS Spam folder
Tokenizer
Wrong Ham
Wrong Spam
Learn Using Spam SMS
User Feed Back

Figure 3 Framework of the spam filtering technique

Figure 3 illustrates the proposed technique that contains analysis engine, tokenizer, stop word filter, dataset, training process, and AIS engine. The following subsections illustrate these components in more detail.

6.1 Analysis Engine

The analysis engine analyzes SMS message to make a reasonable judgment and decision about spamminess. This engine processes data provided by the tokenizer and builds a decision matrix containing the information most relevant to classifying the message. The steps of the analysis engine can be described as follows:

Algorithm 2: Analysis Engine

```
Step1: Load PN, SW, and D from Dataset.
Step2: Build the decision matrix with D, SF (Spam Frequency), HF (Ham Frequency) and ƒ (Affinity).
    Step3: Tokenize SMS (as described in section 6.2).
```

Incoming SMS analyzed by the tokenizer. It examined and divided into smaller components. The analysis engine queries the dataset to identify the importance of each component. Then it calculates the disposition of the message (spam or ham) according to spam score attached with each message.

6.2 The Tokenizer

The tokenizer responsible for breaking the message into colloquial pieces by tokenization process. These pieces can be individual words, or other small chunks of text. The tokenizer starts with separating the message into smaller components, which are usually plain old words. The body and the address parts of a message are parsed, terms are identified based on delimited whitespace and stop marks (e.g. '.', '(', '"', ')', ';', ':', and '-'). Stop words eliminated by stop word filter that will be described in section 6.4. Some other punctuation marks are controversial. Some authors believe that “Free” and “Free?” should be treated the same in most cases as spammy tokens [19].

In the proposed filtering system, the tokenizer instantiate when new training SMS processed. It identifies the different components of the message, so that they can be analyzed by the analysis engine and eventually stored in the dataset. These components are frequently referred to as tokens.

6.3 Stop word filter

The stop word filter gets SMS after tokenization process. It extracts stop words such as (pronouns, prepositions, etc.) and returns the message as a list of keywords. The stop-word filter that has been used in the proposed technique eliminates some words that often occur in many messages (e.g., ‘‘to”, ‘‘a”, ‘‘for”).

6.4 Dataset

A dataset is a catalog of characteristics learned over a period. It provides memory to the other components of the filter and the information necessary to identify the most important characteristics of a user’s SMS.

The proposed dataset divided into three parts: Phone Numbers (PN), Spam Words (SW) and Detectors (D). The proposed system allows the user to enter some mobile phone numbers into PN dataset to block all SMS comes from these numbers. Also, the user has the ability to add some spam words into the SW dataset to block all SMS contains these words. Finally, the proposed system allows the user to train the system using some SMS messages through Detectors (antibodies) dataset that determines if SMS is spam (antigen) or not according to predetermined threshold. As illustrated in Equation 1, all detectors that match the SMS contents are combined to assign a spam score for this SMS.

$$Spam\ Score = \frac{\sum_j^n fi}{NT} \qquad (1)$$

Where, n is the number of matching detectors, NT is the number of message tokens and f_j is the affinity ratio given in Equation 2:

$$f_j = \frac{(Sf)/(TS)}{((Sf)/(TS))+((Hf)/(TH))} \qquad (2)$$

Where, Sf and Hf represent the total number of appearances in spam and ham (non-spam) SMS for the token being computed. TS and TH represent the total number of spam and ham messages.
If spam score is greater than the predetermined threshold value then the SMS considered as spam, otherwise ham.

6.5 The Training Process and User Feedback

The proposed technique used a double collection of bulk SMS messages Spam and Ham to train the filter. For the SMS message classification there exists a problem of finding a reasonable trade-off between two types of errors: classifying legitimate SMS as spam and classifying spam as legitimate SMS. So, the training process of the proposed technique is used to identify spam through the user feedback.

The steps of the training process can be described as follows:

Algorithm 3: User Feedback and Training process

```
Step1: Check SMS Inbox for wrong decision.
Step2: If the classified SMS exists in Inbox (Spam) Then
        2.1 Modify it by "Wrong Decision" (Remove this SMS
        from Inbox and train the system by this message spam)
Step3: If the classified SMS exists in Spam Folder (Ham) Then
        3.1 Modify it by "Wrong Decision"(Remove this SMS
        from Spam Folder and train the system by this message
        as Ham)
```

6.6 The AIS Engine

An immune system’s main goal is to distinguish between self and potentially dangerous non-self elements. In a spam immune system, we want to distinguish legitimate messages (as self) from spam message (as non-self) like biological immune system. The central part of the AIS engine is its Detectors, which are regular expressions made by combining information from training process. These regular expressions match patterns in the entire message. Each Detector acts as an antibody and consists of three associated weights (initialized to zero) detailing what has been matched by that particular Detector [20]:

– Spam Frequency: the cumulative weighted number of spams matched

– Ham Frequency: the cumulative weighted number of messages matched

– Affinity: is a measure that represents the strength of matching between antibody and Message

The AIS engine applies on detectors (antibodies) dataset in two phases. Firstly, it determines the affinity ratio for all detectors with messages; secondly it rejects all detectors with low affinity value, so a clone of detectors with highest affinity is selected. The following algorithm illustrates the steps of the AIS engine:

Algorithm 4: AIS Engine

```
D input: set of Detectors in the Dataset
D' output: set of detectors have a highest affinity capable of classifying SMS.
begin
  Load a set of detectors D from Dataset
  For all  Detectors in D do
    Calculate the affinity for each SMS according to Eq.(2)
  end

  For all Detectors in D do
    Reject the detector with low affinity.
  end
  Select a clone of all detectors that have a highest affinity
End
```

6.7 How the system works

The system has several phases:

- Building the dataset
  - Generation of antibodies from the training process (Algorithm 3)
- Message matching (Algorithm 2):
  - updating of dataset
  - scoring of messages
  - Test message
- Affinity calculation and Expiry of low affinity detectors (Algorithm 4)

These phases represent a cycle that is repeated many times over the life span of the artificial immune system as shown before in figure 3.

Our implementation has been done in Objective-C because of its great flexibility when it comes to working with mobile devices. The detectors (antibodies) library is stored in simple text file. In the library, each line is a regular expression with associated weight. The calculated affinity weight for each message is used to judge if this message is ham or spam according to predetermined threshold.

7. Evaluation Strategy and experimental results

To determine the relative performance of proposed technique, it was necessary to test it against another continuous learning algorithm. The well-known naïve Bayesian classifier was chosen as a suitable comparison algorithm. Naïve Bayesian algorithm is one of the most effective approaches used to classify text documents. Sahami et al. built a Naïve Bayesian classifier for the domain of spam filtering [21]. In this classifier a probabilistic method is used to train a model of classification by using features (keywords) extracted from messages. Give two classes C = {C_1 = spam, C_2 = ham} and features f_1, f_2 , …, f_n the probability that these features belong to a certain class using naive Bayesian can be expressed as follows:

$$P(C|f_1,f_2,\dots,f_n) = \frac{P(f_1,f_2,\dots,f_n|C)P(C)}{P(f_1,f_2,\dots,f_n)}$$

Assuming conditional independence, one can compute $P(f_1,f_2,\dots,f_n|C)$ as follows:

$$P(f_1,f_2,\dots,f_n|C) = \prod_{i=1}^{n} P(f_i|C)$$

To classify an SMS message as spam, one can check if it exceeds a specific threshold as follows:

$$\frac{P(C_1 - spam|f_1,f_2,\dots,f_n)}{P(C_2 = ham|f_1,f_2,\dots,f_n)} \geq \beta, \qquad 0 \leq \beta \leq 1$$

Our main goal was to analyze the detection capability of the proposed technique and naïve Bayesian algorithm on actual SMS messages. We used a collection of English SMS messages, including 1002 legitimate (ham) messages randomly extracted from the NUS SMS Corpus and the Jon Stevenson Corpus, and 322 SMS spam messages collected from the Grumbletext mobile spam site [22, 23]. We divided the corpora into training and testing sets. The training set is the set of SMS messages that gives us a classification result. The test set is actually the SMS messages will run through both the proposed system and Naive Bayesian algorithm which we test to see if classified correctly as spam or not. Initial training was done with 602 ham and 200 spam messages. Then we test the two systems by 1324 SMS messages. We employed the measures that are widely used in SMS spam classification. The common evaluation measures include true positive, true negative, false positive, false negative, detection rate, false positive rate and overall accuracy. Their corresponding definitions are as follows [17]:

True Positives (TP): The number of spam SMS classified as spam.
True Negatives (TN): The number of ham SMS classified as ham.
False Positives (FP): The number of ham SMS falsely classified as Spam.
False Negatives (FN): The number of spam SMS falsely classified as ham.
Detection Rate (DR): TP / (TP + FN).
False positive Rate (FPR): FP/ (TN + FP).
Overall Accuracy (OA): (TP + TN) / (TP + FP + FN + TN).

Table 3 The results (TP, TN, FP, FN, DR, FPR, and OA) of two filtering techniques for different values of threshold

Threshold	Baysian							The Proposed technique						
	TP	**TN**	**FP**	**FN**	**DR**	**FPR**	**OA**	**TP**	**TN**	**FP**	**FN**	**DR**	**FPR**	**OA**
0.25	322	485	517	0	**100%**	**51.6%**	**61%**	322	699	303	0	**100%**	**30.2%**	**77%**
0.3	322	683	319	0	**100%**	**31.8%**	**76%**	322	788	214	0	**100%**	**21.4%**	**84%**
0.35	322	843	159	0	**100%**	**15.9%**	**88%**	322	889	113	0	**100%**	**11.3%**	**91%**
0.4	322	925	77	0	**100%**	**7.7%**	**94%**	322	934	68	0	**100%**	**6.8%**	**95%**
0.45	321	965	37	1	**100%**	**3.7%**	**97%**	320	961	41	2	**99%**	**4.1%**	**97%**
0.5	318	982	20	4	**99%**	**2.0%**	**98%**	314	968	34	8	**98%**	**3.4%**	**97%**
0.55	315	999	3	7	**98%**	**0.3%**	**99%**	303	995	7	19	**94%**	**0.7%**	**98%**
0.6	296	1001	1	26	**92%**	**0.1%**	**98%**	288	1001	1	34	**89%**	**0.1%**	**97%**
0.65	277	1001	1	45	**86%**	**0.1%**	**97%**	265	1001	1	57	**82%**	**0.1%**	**96%**
0.7	253	1001	1	69	**79%**	**0.1%**	**95%**	250	1001	1	72	**78%**	**0.1%**	**94%**
0.75	230	1002	0	92	**71%**	**0.0%**	**93%**	231	1002	0	91	**72%**	**0.0%**	**93%**
0.8	204	1002	0	118	**63%**	**0.0%**	**91%**	221	1002	0	101	**69%**	**0.0%**	**92%**
0.85	94	1002	0	228	**29%**	**0.0%**	**83%**	176	1002	0	146	**55%**	**0.0%**	**89%**
0.9	5	1002	0	317	**2%**	**0.0%**	**76%**	40	1002	0	282	**12%**	**0.0%**	**79%**

The Detection Rate (DR) shows the spam detection accuracy of a classifier. A higher DR indicates better spam detection. The False Alarm Rate (FAR) indicates the false detection of incoming messages. A classifier with a high FAR will move ham messages into the spam folder without user notification. The experimental tests of the two detection algorithms used an iPhone (Apple iOS4) smart phone emulator.

Table 1 gives true positive, true negative, false positive, false negative, detection rate, false positive rate, and overall accuracy with different threshold values of both the Naive Bayesian algorithm and the proposed technique.

Based on Table 3, we can say that (on average) the detection rate, false positive rate, and overall accuracy of the Naive Bayesian algorithm are 80%, 8%, and 89% respectively, while the proposed technique achieved 82%, 6%, and 91% respectively. The performance of the proposed technique is better than Naive Bayesian algorithm.

8. Conclusion

This paper proposed a mobile agent system for detecting SMS-Spam based on AIS. This system contains dataset, tokenizer, analysis engine, stop word filter, AIS engine, and training process. The system used AIS features to building the antibodies (detectors), by initial training phases. The generation, updating, and elimination of detector based on the AIS engine, the content of spam and non-spam SMS Messages used in training. The experimental results applied on 1324 SMS messages show that (on average) the detection rate, false positive rate and overall accuracy of the proposed system are 82%, 6%, and 91% respectively.

References

[1] G. Le Bodic, "Mobile Messaging Technologies and Services SMS, EMS and MMS", 2nd ed., John Wiley & Sons Ltd, (2005).

[2] Mobile SMS Marketing, (December, 2010), available: http://www.mobilesmsmarketing.com/live_examples.php

[3] T. S. Guzella and W. M. Caminhas, "A review of machine learning approaches to Spam filtering", Elsevier, Expert Systems with Applications 36 (2009) 10206–10222

[4] A. Somayaji, S. Hofmeyr, and S. Forrest, “Principles of a Computer Immune System” 1997 New Security Paradigms Workshop, pp. 75–82, 1998.

[5] Healy M, Delany S, Zamolotskikh A., "An assessment of case-based reasoning for short text message classification", In Proceedings of 16th Irish conference on artificial intelligence and cognitive science; 2005. pp 257–66.

[6] Hidalgo JMG, Bringas GC, Sanz EP, Garc FC, "Content based SMS spam filtering", ACM symposium on document engineering. Amsterdam, The Netherlands: ACM Press; 2006.

[7] Gómez, J.M., Cajigas, G., Puertas Sanz, E. Carrero García, "Content Based SMS Spam Filtering", Proceedings of the 2006 ACM Symposium on Document Engineering,

Amsterdam, The Netherlands, ACM Press. Oct., 2006.

[8] He P, Sun Y, Zheng W, Wen X., "Filtering short message spam of group sending using CAPTCHA", In: Workshop on knowledge discovery and data mining; 2008, pp 558–61.

[9] J. W. Yoon, H. Kim and J. H. Huh, "Hybrid spam filtering for mobile communication", Elsevier, computers and security 29 (2010) 446 – 459

[10] S. Sarafijanovic and Jean-Yves Le Boudec. "Artificial Immune System For Collaborative Spam Filtering". In Proceedings of NICSO 2007, The Second Workshop on Nature Inspired Cooperative Strategies for Optimization, Acireale, Italy, November 8-10, 2007.

[11] N. Croft and M. Olivier, "Using an approximated One Time Pad to Secure Short Messaging Service (SMS)", In Proceedings of the Southern African Telecommunication Networks and Applications Conference. South Africa, 2005.

[12] J. Li-Chang Lo, J. Bishop and J. Eloff. "SMSSec: an end-to-end protocol for secure SMS", Computers & Security, 27(5-6):154-167, October 2008.

[13] J. Goodman, G. V. Cormack, and D. Heckerman, "Spam and the ongoing battle for the inbox", Communications of the ACM, 50(2):25–33, 2007.

[14] E. Blanzieri and A. Bryl, " A Survey of Learning-Based Techniques of Email Spam Filtering", Springer Netherlands, Volume 29, Number 1 / March, 2008

[15] J. M. Stewart, E Tittel, and M.Chapple, "Certified Information Systems Security Professional", 4th ed. , Wiley, 2008

[16] S. Dixit, S. Gupta and C. V. Ravishankar, "An Online Detection and Control System for Cellular SMS Spam", IASTED International Conference, Phonex, AZ, USA, 2005

[17] Dong-Her Shih, Ci-Syong Jhuan, Ming-Hung Shih, "A study of mobile SpaSMS filtering system", The XVIII ACME International Conference on pacific RIM management, Canada, July 24-July 26, 2008

[18] K. Rajewsky, "Clonal selection and learning in the antibody system", Nature, 381(6585):751–758, 1996.

[19] J. A. Zdziarski, " Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification ", No Starch Press, 2005

[20] T. Oda and T. White, "Immunity from spam: An analysis of an artificial immune system for junk email detection. Lecture Notes in Computer Science, 3627, 276–289, 2005.

[21] Sahami, M., Dumaisy, S., Heckermany, D., & Horvitz, E.. “A Bayesian approach to filtering junk Emai Technical Report”, Proceedings of AAAI Workshop on Learning for Text Categorization, Madison, Wisconsin, 1998.

[22] SMS Messages for a Public Research Corpus, (May, 2011), available: http://wing.comp.nus.edu.sg:8080/SMSCorpus

[23] Grumbletext, (July, 2011), available: http://www.grumbletext.co.uk