Engineering Applications of Artificial Intelligence 28 (2014) 97–110

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# Hybrid email spam detection model with negative selection algorithm and differential evolution

Ismaila Idris [a], Ali Selamat [a,*], Sigeru Omatu [b]

[a] Software Engineering Research Group (SERG), Knowledge Economy Research Alliance and Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

[b] Department of Electronics, Information and Communication Engineering, 5-16-1 Omiya, Asahiku, Osaka 535-8585, Japan

ARTICLE INFO

*Article history:*
Received 9 April 2013
Received in revised form
16 August 2013
Accepted 2 December 2013
Available online 27 December 2013

*Keywords:*
Negative selection algorithm
Differential evolution
email
Spam
Non-spam
Detector generation

ABSTRACT

Email spam is an increasing problem that not only affects normal users of internet but also causes a major problem for companies and organizations. Earlier techniques have been impaired by the adaptive nature of unsolicited email spam. Inspired by adaptive algorithm, this paper introduces a modified machine learning technique of the human immune system called negative selection algorithm (NSA). A local selection differential evolution (DE) generates detectors at the random detector generation phase of NSA; code named NSA–DE. Local outlier factor (LOF) is implemented as fitness function to maximize the distance of generated spam detectors from the non-spam space. The problem of overlapping detectors is also solved by calculating the minimum and maximum distance of two overlapped detectors in the spam space. From the experiments, the results show that the detection accuracy of NSA–DE is 83.06% while the standard negative selection algorithm is 68.86% at 7000 generated detectors.



## 1. Introduction

The cheapest and most important form of communication in the world today is email. It is effective, simple and available for all computer users. The simplicity of email makes it vulnerable to a lot of threats. One of the most important threat to email is spam; virtually all email users across the world suffer from email spam (Cormack et al., 2011). The word spam was used to describe unwanted, junk mails sent to an internet user's inbox. It is very convenient for spammers to send millions of email spam all over the world with no cost at all (Carpinter and Hunt, 2006). This makes it a common scenario for all internet users to receive junk mail hundred times daily. Different techniques have been adopted to stop the threat of spam or drastically reduce the amount of spam that attacks internet users across the world. An anti-spam law was enacted by legislating penalty for spammers that distribute email spam (Schryen, 2007). Also, two general approaches have been used in email spam detection; a knowledge engineering approach and a machine learning approach (Wamli et al., 2009). In the knowledge engineering approach, the use of network information and internet protocol address techniques to determine if a message is spam or non-spam is called origin-based filter. Sets of rules have to be specified in the knowledge engineering approach in order to determine which email is to be categorized as spam or non-spam. Such rules could be created by the use of filter or by some other authority. An example of this process is the software company that provides a particular rule based spam filtering tools. By the application of this method, there is promising result. However, the rules need to be maintained all the time and updated which is a waste of time and inconvenient for most users. Machine learning is more efficient than knowledge engineering approach (Guzella and Caminhas, 2009) and does not require specifying rules; a set of pre-classified email message (training sample) is applied. Specific algorithms are used to learn the classification rules from the email messages. The filtering techniques are the most commonly used methods; it identifies whether a message is spam or non-spam based solely on the message content and some other characteristics of the message. Despite different approaches and techniques adopted to fight the scourge called spam, the internet today still witnesses huge amount of spam (Zhang et al., 2004; Massey et al., 2003), and more attention is needed by adaptive techniques on how the menace can be drastically reduced if not totally eliminated.

Due to the wide knowledge of machine learning approach, several algorithms have been used for email spam detection (Guzella and Caminhas, 2009). They include artificial immune system (AIS), support vector machine (SVM), neural network

* Corresponding author.
*E-mail addresses:* ismi_idris@yahoo.co.uk (I. Idris), aselamat@utm.my (A. Selamat).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.12.001

(NN), Naïve Bayes (NB), *k*-nearest neighbour (KNN), etc. In this paper, we propose a new approach that is inspired by artificial immune system model; that is a negative selection algorithm (NSA) with the combined effort of differential evolution (DE) which modifies the standard negative selection algorithm in order to generate more accurate results. The engineering goals required in hybrid negative selection algorithm can be viewed in three ways; first, is to generate an efficient detector set; secondly, is to limit the number of detectors that will be generated and thirdly, is to maximize the detector set distance as much as possible. Problems that require attention in this research work are: (i) generating detectors in the spam space; (ii) maximizing distance between spam detectors and the non-spam space and (iii) solving the problem of overlapping detectors in the spam space. These problems are solved by the implementation of local differential evolution for generating detectors, application of local outlier factor as fitness function to maximize the distance between generated detector in the spam space and the non-spam space, calculating the minimum and maximum distance between two overlapped generated detectors as fitness function. The performance of NSA is determined by detector generation and how effective it is able to utilize the detector coverage space of spam and non-spam. This paper is organized into six sections, Section 1 is the introduction, Section 2 discusses the related work in negative selection algorithm, the proposed improved model and its constituent framework are presented in Section 3. Empirical studies, results and discussions are presented in Section 4, Section 5 discusses the experimental results while conclusions and recommendations are presented in Section 6.

## 2. Related work

Over the past years, rapid expansion of computer network systems has changed the world. The expansion is essential for an effective computer security system because attacks and criminal intend are increasingly popular in computer network (Golovko et al., 2010). Negative selection algorithm, while not reacting to the self cells uses the immune system capability to detect unknown antigens. Its mechanism protects body against—reactive lymphocytes. Receptors are made through a pseudo-random genetic rearrangement process during the generation of T-cells (Wang and Zhao, 2008); they then undergo a censoring process in the thymus called the negative selection. In this process T-cells that do not bind to self-proteins are destroyed. Therefore, immunological function and protection of the body against foreign antigens is possible through circulation of matured T-cells (Zhang et al., 2010). Recent work uses immunological function to solve complex problems in negative selection algorithm. The work of Gong et al. (2012) introduced a further training strategy to generate more self-detectors to be able to cover the self-space for effective detectors. The technique reduces the false rate, as wrongly classified non-self will be re-classified for correctness. The drawback of the techniques was that it leads to scalability and part of the self-detector may be covered by overlapped detectors. An immune local concentration based detection approach was proposed by Wie et al. (2011), two element local concentration as a feature vector was combined with negative selection algorithm and optimized with genetic algorithm. The technique generates effective and efficient detectors as the local concentration feature vectors are optimized before training the features. The technique is computationally expensive and also time consuming in achieving desired feature detectors. A similar work was presented by Yildiz (2009, 2013a) and Prakash et al. (2008) to implement optimized immunological functions in solving complex problems in industry. This technique uses evolutionary algorithm to implement parameters optimization in the immune system. A detection model based on penalty factor was proposed by Zhang et al. (2010) to construct a model for spam detection; by redefining the harmfulness of self and non-self using the negative selection algorithm penalty factor to divide the candidate signature library into two detection signature libraries as a self-detector, and then splitting of the programs in an orderly way into various short bit strings. The work of Xin et al. (2010) and Yuebing et al. (2010) proposed a self-detector by the use of real valued negative selection algorithm. A variable size *r*-contiguous matching rule was implemented and the value of the variable size *r* was used to balance between more generalization and specification of the self-space. This technique is not very sufficient in generating self-detectors as it has a constant threshold value which may lead to over-fitting problems in most cases. A shape space as an occupancy of negative selection algorithm was proposed by Wanli et al. (2010). The work states the importance of full coverage of the shape space for effective detectors by suggesting a heuristic for detection generation which was demonstrated by antigen feedback mechanism. The issue of overlapping and scalability was not addressed by Wanli et al. (2010); it will definitely have effect on the shape space as effective detectors generated are unable to sufficiently cover the shape space. The work of Forrest and Perelson (1994) quantifies the number of resources that will be required by NSA in order to exhibit a very good detector capability rate and failure rate. Forrest and Perelson (1994) use a single global affinity threshold value *r* which ranges between a specific number with *r*-contiguous bits matching rule for each and every instance within its population. The affinity threshold in this case is determined through a trial and error method, where the best threshold with the best performance is targeted as the affinity threshold.

The understanding of artificial immune system (AIS) based on the mammalian immune system is vital for this study. A comprehensive artificial immune system survey was analysed by Dasgupta et al. (2011). The research discusses the history, recent development and the four major AIS algorithms. The main goal of the immune system is to distinguish between non-self and self element which is the basis for our implementation with negative selection algorithm (NSA), one amongst the algorithm of artificial immune system (AIS). This research will replace self in the mammalian immune system as non-spam in our system and non-self in the mammalian immune system as spam in our system. Artificial immune system (AIS) is a new mechanism implemented in the control of email spam. Pattern matching was used to represent detectors as regular expression by Oda and White (2003) in the analysis of message. A weight is assigned to the detector which is decremented or incremented when observing the expression in spam message with the classification of the message based on threshold sum of the weight of matching detectors. The system is meant to be corrected by either increasing or decreasing of all matching detector weights with 1000 detectors generated from spam-assassin heuristic and personal corpus. The results were acceptable based on small number of detectors that was used. A comparison of two techniques to determine message classification using spam-assassin corpus with 100 detectors was proposed by Oda and White (2003). This approach is like the previous techniques but the difference is the increment of weight where there is recognition of pattern in spam messages. Random generation of detectors does not help in solving the problem of best selected features; though, feature weights are updated during the matching process. The weighting of features complicates the performance of the matching process. More experiments are performed by Oda and White (2005) with the use of spam-assassin corpus and Bayesian combination of detector weights. Messages are scored by simple sum of the message matched by each non-spam in the detector space and also the use