Negative Selection Algorithm In Artificial Immune System For Spam Detection

Ismaila Idris
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia,
Johor Darul Ta'zim, Malaysia.
ismi_idris@yahoo.co.uk.

Ali Selamat
Faculty of Computer Science and Information System,
Universiti Teknologi Malaysia,
Johor Darul Ta'zim, Malaysia.
Selamat.ali@gmail.com

Abstract- Artificial immune system creates techniques that aim at developing immune based models. This was done by distinguishing self from non-self. Mathematical analysis exposed the computation and experimental description of the method and how it is applied to spam detection. This paper looked at evaluation and accuracy in spam detection within the negative selection algorithm. Preliminary result or classifier of self and non-self was carefully studied against mistake of assumption during email classification whereby an email was recognized as a spam and deleted or non-spam and accepted carelessly. This process is called false positive and false negative. Given a threshold, the accuracy increase with increased threshold to determine best performance of the spam detector. Also an improvement of the false positive rate was determined for better spam detector.

Keywords- *Artificial immune system; Negative selection; Computer security; Algorithm. Model.*

I. INTRODUCTION

Over the past years, rapid expansion of computer network system as change the world. It is essential for an effective computer security system because attacks and criminal intend are increasingly popular in computer network[1] . Negative selection algorithm, will not react to the self cells uses the immune system capability to detect unknown antigens. Its mechanism protects body against self reactive lymphocytes. Receptors are made through a pseudo-random genetic re-arrangement process during the generation of T-cells [2]; 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 [4].

We shall first of all generate a spam and non-spam detector. E-mail classification will then take place by utilizing the non-spam and the spam accordingly in other to successfully reduce the false rate. Our improved classification techniques are also compared with the existing techniques. The experiment confirms the reliability and efficiency of our techniques in minimizing false rate. The datasets used in this research is gotten from machine learning repository, Center for Machine Learning and Intelligent System.

II. COMPUTATIONAL BASED SECURITY SYSTEM.

The ability of immune system adapting to spam is also applied in computer concept, most often in issues or problem of recognizing new spam intruding the system and detecting an intrusion into a network. The algorithm is on the bases of creating T-cell in the thymus. Self is been describe as the normal state of the computer system and training detectors on the normal state of the computer to produce detectors that would recognize invaders as abnormalities known as the non-self in the system. Since training data set was carried on self (self detectors) therefore rejection of match data set with self occur and only dataset that did not match with self detector will exist in the valid detector set to form a detector

Figure.1. below generates a valid detector set and new instance for computer spam detection.

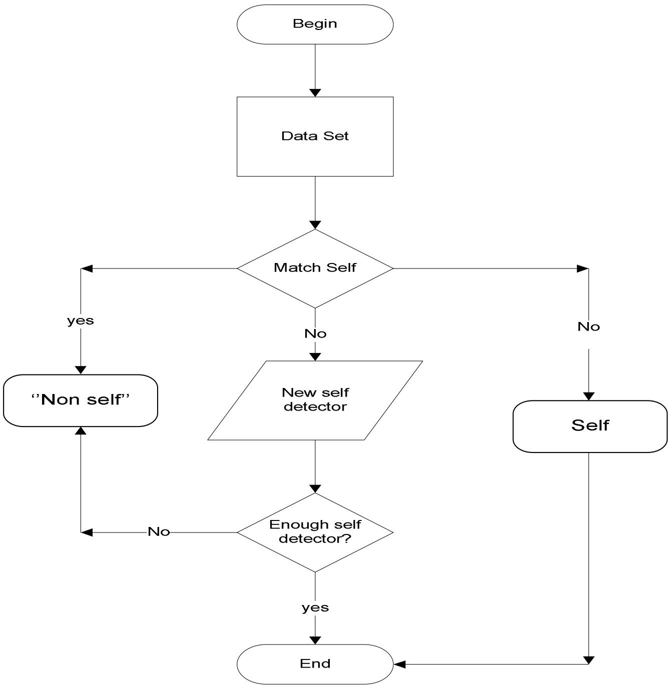

Fig 1. Valid detector and detector of new instance

III. NEGATIVE SELECTION BASED ALGORITHM

The negative selection algorithm defines self to be equal to the collection of element in a feature space U.
U is represented by list of features which corresponds to the space of a system.
where S=Subset of space that are considered as normal for the system.

R= Set of detectors generated. (1)

$R = S$ R fails to match any string in S (2)

This approach analyzes happenings in Negative immune system by generating random detectors and discard those that match any element in the self set [3]. Continually matching S for changes with detectors R against S if S ever matches R. Change is known to have occurred as detectors are not suppose to match any string in S.

The algorithm is as stated below.

```
DEFINITIONS:
x is a self data set (spam)
y is a non-self data set (non-spam)
N is the number of matching data
SM(0)=0, NSM(0)=0;
INPUT:
        ∝ /* ∝ is a threshold
        b /* b is the detector of x;
        a /* a is the detector of y;
OUTPUT:
        Finding matching detector of both self and non-self
BEGIN
Input N;
Input Sm(1), Nsm(1) /*Sm is self matching and Nsm is non-self matching;
For i=1 to N
    Sm(i) = Sm(i) + Sm (1-- i);
Next;
For i =1 to N
Nsm (i) = Nsm (i) + Nsm (i - 1);
If faffinity>= ∝
        f affinity (x) = max;
        f affinity (y) = max;
end if
if fmatching = =.T.
        (b,x) >= ∝;
else
        (a,y) >= ∝;
End if
End
```

Fig.2. Negative Selection Algorithm.

IV. E-MAIL CLASSIFICATION

In the process of e-mail classification, two mistakes occur. It is either the e-mail is recognized as self were as it is non-self and then it is deleted mistakenly or been mistaken as non-self instead of self and it is accepted carelessly. This process is called false positive and false negative [6].
The false positive occurs when the email or data that are needed to create a detector are classified as self while emails or data that are supposed to be discarded are recognized as non-self. This scenario (false negative and false positive) is calculated using classification accuracy which is the main measure of performance. Classification accuracy deals with false positive, false negative and accuracy whose formulae are used to compare different classifier performance [4]. False positive is the percentage of non-self data classified as self while false negative is the percentage of self data which are classified as non-self and accuracy is calculated by the formulae below.

Accuracy = ((TP+TN)/(S+NS)) X 100 (3)

Where TP represent True Positive; TN represent True Negative; S represent Spam and NS represent Non Spam [7].
The figure below demonstrates how false positive and false negative are calculated. The first row depict the total non-spam that is divided to true positive and false positive. These rows contain total dataset which are non spam and some are wrongly classified as spam (FP) while others are assigned correctly as non-spam while the opposite is the case with the second row.

Non-spam	spam
True positive (TP)	False positive (FP)
False negative (FN)	True negative (TN)

Fig. 3. False positive and False negative

None of the anti spam solution that has been proposed on false positive and false negative approach perfection [5]. Though the result of spam is reduced but not completely.

V. PROBLEM DEFINITION

Statistical model is the most common approach to spam detection in computer security [6, 7]. The spam probability is higher in calculating the probability of occurrence of a given value. In other methods, models are built to predict the future behavior of systems or processes base on recent and formal states [8, 9]. In this scenario, a spam alarm is raised if the normal states of the system differ from the predicted state.
Generally, spam is considered as a deviation from a set of normal states with assumption of distance in this space that allows to measure for deviation [10].
In identifying the state of a system as self or non self, is the reason for spam detection problem definition. This is represented by a set of features as shown below:

Let's make vector of the feature as the set of the system which is the system state space.

$$a^{i} = (a_{1}^{i}\text{---------------}a_{b}^{i}) \epsilon\ [0,1]^{b} \qquad (4)$$

Each state of the feature been represented by a set

$$\cup\subseteq [0,1]^{b} \qquad (5)$$

This includes the feature vectors which correspond to all the probability state of the system.

$\cup$ = Universal set of all the system. (6)

0 and 1 represent system being a self or non-self. (7)

For normal subspace (crisp characterization) set of feature vectors self$\subseteq\cup$, which indicate a non-spam.
Therefore non-self which is its compliment defined as non-self = u – self, where non-self indicate spam in the system.
Using its characteristics function, we can define self or non-self as follows

$$a_{self}\colon [0,1]^{b} \longrightarrow [0,1] \qquad (8)$$

$$a_{self}\colon \overrightarrow{(a)} \begin{cases} 1\ if\ \vec{a}\in self \\ 0\ if\ \vec{a}\in non-self \end{cases} \qquad (9)$$

The above is characterized by artificial immune system were cells are distinguish from foreign antigens. The term self represent cells in the immune system and non-self represent the foreign antigens in the system. Though, there are difference between the spam and the non spam state.
Where as normal subspace (non-crisp characterization) features of a spam and non spam is extended to pick values of intervals [0,1]

$$\beta^{self} : [0,1]^{a} \rightarrow [0,1] \qquad (10)$$

This value represent degree of either it is a spam or not a spam. Where by 1 indicates that it is not a spam and 0 indicates that it is a spam. The intermediate value represent element with some degree of being a spam or non spam.
Also, binary decision as to be in cooperated. It is simple to go from non-crisp characterization to the crisp one by creating a limitation.

$$\beta self, \mathrm{L}\ (\vec{a}) = \begin{cases} 1\ if\ \beta\ self\ (\vec{a})>L \\ 0\ if\ \beta self\,(\vec{a})\leq L \end{cases} \qquad (11)$$

With the sample $self^{I}$ ⊆ self, a good estimate of the normal space can be created with characteristic function. a^{self} in the crisp characterization case and $\beta\ self$ In the non crisp characterization case. This function should be able to tell if there is spam or not.

VI. EXPERIMENT AND RESULTS

This paper looks at algorithm of spam detection using negative selection based on classification of spam content on a network thereby increasing the accuracy of spam detection with negative selection techniques. The data set used in this technique has 4601 instances in which 39.4% are spasms and each instances has 57 attributes. The data set are loaded and divided in to two classes. The training data set which is meant for training the data and the testing data set which is meant to test the trained data set. The spam e-mail are called spam corpus and are separated from the training data set. The spam corpus is divided in to exemplar and training data set. Let's assume that about 50% of the spam data are exemplars, the exemplar is initialized from random sample of trained data set. The distance between the exemplar classes and training data sets are calculated by Euclidean distance formulae which is use for data classification. The minimum distance is selected to determine the class of exemplar for classification. The Euclidean formula for continuous value is stated as below.

$$\text{Total Distance (X1,X2)} = \sqrt{\textstyle\sum_{i=1}^{n}(X1i - X2i)} \qquad (12)$$

X1, X2 and I are the class and specified attribute of exemplar training data.
Matching is applied on the chosen classes of exemplar after initialization. Total rate is the bases for selecting the classes. The percentage of classes which are selected from exemplar randomly is the total rate. Threshold rate is the percentage of bits from each class which are selected for matching; the bits of class are selected base on threshold rate of matching. Classification performance is estimated by Euclidean distance formulae at matching class in every iteration. The matching class becomes the new class with better performance. This becomes the new detector with better performance.
The distance between the optimized detector and training data set are finally estimated and the class with minimum distance value is selected for classification.
Performance measurements using the proposed technique were presented in the figure below with different threshold for testing data.

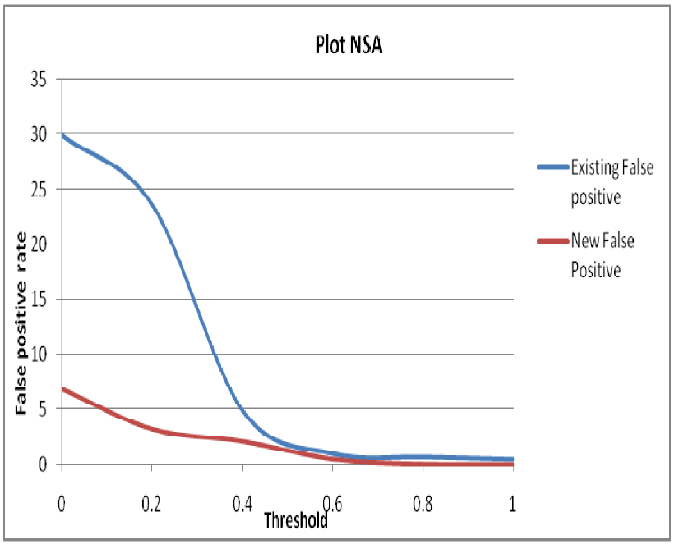

Fig.4. Existing False rate and new false rate

Fig.4. analyzes the false positive rate of both existing model and our improved model with reduced false positive of our proposed model with threshold of between 0 and 1. Accuracy increased at best in the proposed model by 94.28% , worst at 94.05% compare with the existing model whose accuracy with same threshold was at 89.3% best and 82.9% worst. This indicates an increase in the accuracy performance of the proposed model. Also, false positive rate decreased at threshold 1 in the proposed model by 0.01% at its best and 6.8% worst while the existing model whose false positive rate with same threshold was 0.5% at its best and 29.9% at its worst. This shows an improvement of the false positive rate of the proposed model.

VII CONCLUSION

Our network security has been improved due to assistance of Artificial immune systems where computer system complexity is fast becoming a worrying issue and as tremendous influence in spam propagation. Antivirus finds it difficult to detect spam these days as it has become invisible in our computer system. In this paper, we presented the self and non-self in a way to create efficiency of detectors. The novelty of this paper was to look in to false rate with respect to self and non-self by reducing the false rate for effective detector. Subsequent research shall be looking at constant upgrade of the existing model of antibody (self) in other to prepare it against new spam.

REFERENCES

1. Golovko, V., et al., Neural network and artificial immune systems for malware and network intrusion detection, J. Koronacki, et al., Editors. p. 485-513. 2010.
2. Wang, C. and Y. Zhao, A new fault detection method based on artificial immune systems. Asia-Pacific Journal of Chemical Engineering,. **3**(6): p. 706-711. 2008.
3. G6mez', J., F. GonzAlez, and D. Dasgupta, An Immuno-Fuzzy Approach to Anomaly Detection. The IEEE International Conference on F uzz y Systems 2003.
4. Zhang, Y., et al., Immunity-based model for malicious code detection. Changsha. p. 399-406. 2010:
5. Minh Tran and G. Armitage, Evaluating The Use of Spam-triggered TCP/IP Rate Control To Protect SMTP Servers. Australian Telecommunications Networks & Applications Conference 2004 (ATNAC2004), Sydney, Australia. , December 8-10 2004.
6. Eskin, E., anomally detection over noisy data using learned probability distribution. In Proceedings of the International Conference on Machine Learning.2002
7. Denning, D.E., an intrusion detection model. IEEE,:p. 118-131. 1986
8. Wang, Q. and X.K. Feng. A detector generation algorithm based on negative selection. 4th International Conference on Natural Computation, ICNC, Jinan. vol.6 pp: 605-611 2008
9. Terran Lane and C.E. Brodley,Temporal Sequence Learning and Data. ACM Transactions on Information and System Security, . **Vol. 2, No. 3,**: p. Pages 295–331. August 1999.
10. Esponda, F., S. Forrest, and P. Helman, A Formal Framework for Positive and Negative Detection Schemes. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,. vol.**34. no.**1: p. 357-373. 2004