AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

Email Spam Detection Generation Algorithm for Negative Selection Algorithm with Hamming Distance Partial Matching Rules

[1]Ismaila Idris, [2]Shafi'i Muhammad Abdulhamid, [3]Abdulmalik Mohammed, [4]Umar Dauda Suleiman, [5]Nicholas Akosu

[1,2]Department of Cyber Security Science, Federal University of Technology, P.M.B 65, Minna, Niger State, Nigeria.
[3]Department of Computer Science, Federal University of Technology, P.M.B 65, Minna, Niger State, Nigeria.
[4]Department of Electrical and Electronics Engineering, Federal University of Technology, P.M.B 65, Minna, Niger State, Nigeria.
[5]Department of Computer Science, Federal Polytechnic Nasarawa, Nasarawa State, Nigeria.

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form
8 April 2014
Accepted 20 April 2014
Available online 10 May 2014

Keywords:
Negative selection algorithm, time, accuracy, detector generation, r-chunk, hamming.

ABSTRACT

Negative selection algorithms (NSAs) are inspired by artificial immune system. It creates techniques that aim at developing the immune based model. This is done by distinguishing self from non-self spam in the generation of detectors. In general, NSAs has an exponential run time. This research study the significance of time and accuracy for two commonly used matching rules. The hamming and r-chunk matching rules, based on different threshold values (r) for generating set of fixed number of detectors. The results show the differences between the mean values of time and accuracy for hamming and r-chunk matching rules. Statistical t test shows that the difference between hamming and r-chunk matching rule are insignificant for accuracy while it is significant for time.

To Cite This Article: Ismaila Idris, Shafi'i Muhammad Abdulhamid, Abdulmalik Mohammed, Umar Dauda Suleiman, Nicholas Akosu., Email Spam Detection Generation Algorithm for Negative Selection Algorithm with Hamming Distance Partial Matching Rules. ***Aust. J. Basic & Appl. Sci.,*** *8(6): 21-29, 2014*

INTRODUCTION

Negative selection algorithm (NSA) has been used successfully for a broad range of application in the construction of artificial immune system (Balthrop, J., *et al*., 2002). The algorithm is an approach that deals with anomaly detection by the use of negative selection which was originally proposed by Forest *et al* (1994). 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 (Wang, C. and Y. Zhao, 2008); The immature T-cell undergo both positive selection and negative selection procedure in the thymus. The NSA was inspired by the negative selection techniques that exist within the natural immune system (NIS) (Cantu-Ortiz, F.J., 2014; Travé-Massuyès, L., 2014). In this process, T-cell that those not bind to self protein are destroyed. In a nutshell, the immunological function and protection of the body against foreign antigens is possible through circulation of matured T-cells (Zhang, Y., *et al*., 2010). The main concept behind the negative selection algorithm is to generate sets of candidates, c such that $\nvdash x_{i\epsilon} \in s\ f_{MATCH}\left(x_i, z_p\right) < r$. The concept is illustrated in figure 1. This paper considered negative selection algorithm in binary classification operating on a string space. Classification is one of the familiar techniques used in machine learning. Patterns belonging to different classes are discriminated due to the generation of decision boundaries.

Single global affinity threshold, r, was originally used by forest et al with the r- contiguous matching rule for each individual artificial lymphocytes (ALC) around the population of the ALCs, C. A process of trial and error is use to determine the affinity threshold, then, the threshold that yields the best system performance will be chosen as the targeted threshold for the system. A frame work that will aid in the chosen of an optimum value for r in conjunction of r-contiguous rule was provided by (Ayara, M., *et al*., 2002). Previous research work that adopts the partial matching rule with hamming distance shows no efficient detector generated (Haiyu Hou and G. Dozier, 2006). This paper analyzes run time for generating efficient detector algorithm that adopt the hamming distance matching rule while comparing its performance with the r-chunk matching rule.

Section 2 presents related work and the contribution of this paper while section 3 discuss the algorithm of the r-chunk matching rule and section 4 discusses the algorithm of the hamming distance matching rules. In

Corresponding Author: Ismaila Idris, Department of Cyber Security Science, Federal University of Technology, P.M.B 65, Minna, Niger State, Nigeria.
E-mail: ismaila.idris95@gmail.com.