# ACADEMIC LEADERSHIP BIO-INSPIRED CLASSIFICATION MODEL USING NEGATIVE SELECTION ALGORITHM

**Hamidah Jantan, Siti 'Aisyah Sa'dan, and Nur Hamizah Syafiqah Che Azemi**

*Universiti Teknologi MARA [Terengganu], Malaysia, hamidahjtn@tganu.uitm.edu.my*
*Universiti Teknologi MARA [Terengganu], Malaysia, sitiaisyah@tganu.uitm.edu.my*
*Universiti Teknologi MARA [Terengganu], Malaysia, hamizah_syafiqah@yahoo.com*

**ABSTRACT**. Negative selection algorithm has been successfully used in several purposes such as in fault detection, data integrity protection, virus detection and etc. due to the unique ability in self-recognition by classifying self or non-self's detectors. Managing employee's competency is considered as the top challenge for human resource professional especially in the process to determine the right person for the right job that is based on their competency. As an alternative approach, this article attempts to propose academic leadership bio-inspired classification model using negative selection algorithm to handle this issue. This study consists of three phases; data preparation, model development and model analysis. In the experimental phase, academic leadership competency data were collected from a selected higher learning institution as training data-set based on 10-fold cross validation. Several experiments were carried out by using different set of training and testing data-sets to evaluate the accuracy of the proposed model. As a result, the accuracy of the proposed model is considered excellent for academic leadership classification. For future work, in order to enhance the proposed bio-inspired classification model, a comparative study should be conducted using other established artificial immune system classification algorithms i.e. clonal selection and artificial immune network.

**Keywords**: bio-inspired algorithm, negative selection algorithm, classification, academic leadership

## INTRODUCTION

The negative selection algorithm is one of the most widely used techniques of the immunological principles used in artificial immune system. This approach is developed as a new paradigm in computational intelligence inspired by the biological immune system for data analysis (Timmis, Neal, & Hunt, 2000). The first negative selection algorithm was proposed by Forrest in identifying information affected by the biological based infection that was transformed to machine learning framework (Forrest, Perelson, Allen, & Cherukuri., 1994). It is primarily used to detect changes in pattern's behavior by gene detectors in the complementary space. In the original version of the negative selection algorithm, the detectors are used directly to classify new data as self (normal) or non-self (abnormal). Due to the uniqueness ability in self-recognition, this algorithm has been mainly used for anomaly detection by classifying self or non-self as detectors in network security, computer virus detection, network intrusion detection, fault diagnose and many others(We, Zheng, & Wang, 2004). Besides that, negative

selection algorithm can also be used for classification by matching the self or non-self's detectors as a recognition model in many areas. Due to that reason, this paper attempts to study the potential of this algorithm for talent recognition in Human Resource field especially in the talent management.

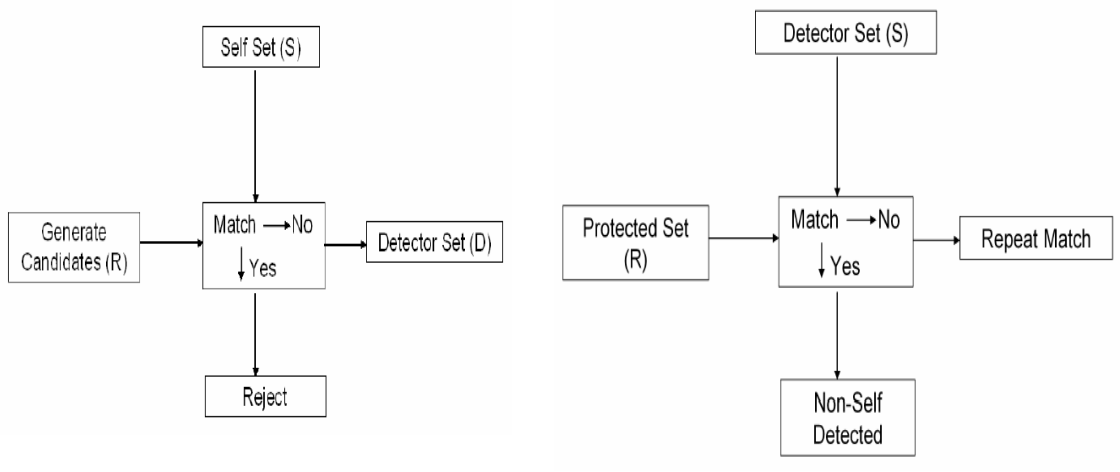
In Human Resource (HR) field, the process of talent recognition involves a lot of managerial or human decisions which are very subjective, uncertain and difficult especially in managing employee's advancement through promotion process (Jantan, Hamdan, & Othman, 2011). The employee's recognition is a way for an employee to improve his/her career path development and it depends on the involvement of several approaches and employee's competency criteria. In higher learning institution, competency criteria for academic leadership are related to his/her ability in several areas such as teaching and supervision, research and publication, contribution to university or society and many others. In this study, negative selection algorithm is proposed as the recognition mechanism in identifying a potential employee for a suitable position based on biological immune system approach. The rest of this paper is organized as follows: the second section discusses the related work and study motivation in regards to negative selection algorithm and academic leadership in higher learning institution, the third section describes the experiment setup conducted in this study and, the fourth section discusses the results and discussions. Finally, the paper ends with the fifth section where the concluding remarks and future research directions are identified.

## RELATED WORK

### Negative Selection Algorithm

Negative Selection (NS) algorithm is known as classification algorithm that mimics or simulates the process of negative selection in the vertebrate immune system. The negative selection is one of the mechanisms in the natural immune system that has inspired the developments of most of the existing Artificial Immune systems. In the T-cell maturation process of the immune system, if a T-cell in thymus recognizes any self-cell, it is eliminated before deploying for immune functionality. Similarly, the NS algorithm generates detector set by eliminating any detector candidate that matches elements from a group of self-samples (Mohamed, Ammar, & Rajasekaran, 2012). A population of detectors is created to perform the job of T-cells. These detectors represents fixed length binary strings and a simple rule is used to compare bits in two such strings, followed by the decision on whether matching has occurred or not. Such matching process is equivalent to a match between lymphocyte and antigen. They are trained on unlabeled data sampled from a certain sub-region of the problem domain, and then used to determine whether or not new unseen data points belong to the same sub-region.

Besides that, NS algorithms are based on so-called detectors, which can be understood as patterns that match to small subsets of the problem domain. In NS algorithm, the first step is to generate a set of detectors. Each detector is a string that does not match a predetermined substring of the protected data. For matching, usually a partial matching rule is defined, because it can be extremely rare that random strings that are generated exactly match the source data, even if these strings are small. The second step is to continually monitor the data by comparing them with the detectors. If a detector is ever activated, a change is known to have occurred. Although this approach might seem too simple to work, it is rather effective: a fairly small set of detector strings has a very high probability of noticing a random change to the original data (Ji & Dasgupta, 2007). The NS algorithm basically consists of two steps as shown in Figure 1, where the integrity of a data file or string has to be protected (Soam, Khan, Bhasker, & Mishra, 2011).




(a) Generation of Detector Set (b) Detection of New Instances

**Figure 1. Typical Negative Selection Algorithm**

NS algorithm has been applied in many areas, mainly for anomaly detection by classifying self or non-self as detectors in network security, computer virus detection, network intrusion detection, fault diagnose and etc.(We et al., 2004). However, there are many other areas, such as manufacturing, engineering, medicine, finance, human resource and others, which should consider this algorithm as the potential approach for knowledge analysis and discovery.

**Academic Leadership in Higher Learning**

The fundamental form of academic leadership is 'intellectual leadership'. This is the development of leading ideas and the formation of new academic directions. Most academics aspire to this. Necessarily, given the narrow and compartmentalized forms of knowledge in academia, such leadership has a limited scope, at least within a single university, though it may embrace a broader subject community. A university needs to be able to recognize and reward this but cannot harness it much beyond the immediate context in which it arises. In university, the academics are expected to teach, supervise research, publish scholarly works, conduct consultancy, and provide service to community. These duties need to be considered as academic talent criteria in managing academic leadership for higher learning institution (Tarique & Schuler, 2010). Nonetheless, these academic contribution aspects also depend on several factors, such as universities' direction, academic appointment, seniority, and specialization (Ismail & Rasdi, 2008).

In universities, teaching, supervision and research activities are known as academics main duties towards academic leadership development. Most of academic research is comprised from the individuals who are doing academic research which can be extreme, inspiring, and worthwhile. At the same time it is important to realize that a research profession includes numerous exercises other than exploration. Researchers invest their time composing for financing their research, and proposing experimental papers to report research discoveries in their findings. Moreover, they also devote their time presenting their findings in institutional discussions, and scientific conferences. A researcher's life is commonly loaded with tasks that need to be carried out and most of the researchers work hard since they love what they do (Vincent-Lancrin, 2006). All these aspects should be taking into consideration in evaluating the competency of academic leadership in higher learning institution as an academic talent.

Besides that, academic talent marketplace is highly competitive at international level which is considered as current and future direction for any education-based institution (Verhaegen, 2006). Academic talent that is measured by their academic leadership aspects is a valuable asset for higher learning institution in order to ensure the development of excellence as their core business. In higher learning institution, this issue associated with recruitment and retirement of academic talent as the key of long-term success and competitiveness. Therefore, this study finds the importance of classifying academic talent through academic leadership criterion. Nowadays, there are studies done on this issue which applied soft computing and data mining techniques (Chang, Cheng, & Chen, 2007; Jantan, Hamdan, & Othman, 2009). However, evolutionary computation bio-inspired algorithm, such as Genetic Algorithm, Ant Colony Optimization, Artificial Immune System (AIS) and many others, has not attracted researchers in this area. Due to this reason, this study attempts to discover the potential of using bio-inspired algorithm to deal with this issue.

## EXPERIMENT SETUP

Negative selection classification process has two phases; the first phase is learning process whereby training data were analysed. The second phase is classification process, where testing datasets were used to estimate the accuracy of the propose classification model. If the accuracy is acceptable, the model can be applied to new data (untrained data) for classification. In the experimental phase, this study was aimed to discover employee's leadership bio-inspired classification model using negative selection algorithm. There were three phases involved i.e. gathering information and data preparation; classification model development; and model analysis. The first phase began by collecting and conducting data pre-processing for data obtained from academic promotion evaluation. The collected data contain demographic information and evaluation marks for seven criteria for academic promotion, as shown in Table 1.

**Table 1. Attribute Description**

| Criteria | Attributes |
|---|---|
| Demographic | Year<br>Gender<br>Grade Promotion |
| Academic Leadership Criteria | Teaching and supervision<br>Research and publication<br>Consultation and/or Expertise<br>Conference Participation<br>Service University/Community<br>Academic award<br>Leadership and personal attitude |

In the model development phase, the main idea of NS algorithm is to generate a set of detectors by first randomly making candidates, followed by discarding those that recognize training self-data, and later, these detectors were used to detect anomaly or recognize self or non-self's detector for classifying of a new candidate. NS algorithm consists of three phases: defining self-data, generating candidate detector and matching the generated detector with self-data based on affinity threshold. The threshold value is a value that can be a benchmark to distinguish between the normal and the abnormal data in the data set. It is important because it helps to filter the abnormal data to be accepted as non-self's data. The threshold value was carefully chosen based on the value that can produce the similar output as in the real situ-

ation. This matching was intended to know which data were similar to the data in Memory. The matching will be known by calculating their affinity measure. The affinity measure is calculated by using Euclidean distance formula. The Euclidean distance is a formula that calculates the remove between two focus points as the square root of the total of the squares of the contrasts between the comparing directions of the focuses.

The second phase is the model development phase which consists of two processes, which are training phase and testing phase, as shown in Figure 2.

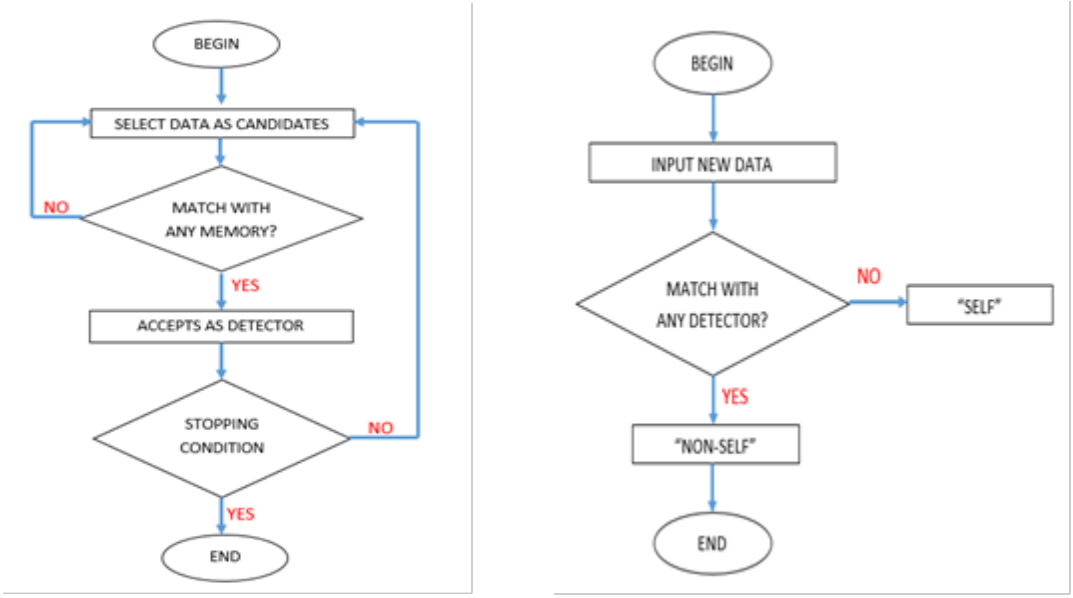



**Figure 2. NS Classifcation Model Development**

The sample of datasets for training and testing is shown in Figure 3 and the list of detectors is shown in Figure 4.

```
trainingData[52]  100.00,57.33,45.00,60.00,53.00,30.00,75.00
trainingData[53]  100.00,25.00,20.00,40.00,59.00,50.00,70.00
trainingData[54]  100.00,20.00,50.00,28.00,55.00,70.00,95.00
trainingData[55]  100.00,48.50,70.00,50.00,50.00,40.00,70.00
trainingData[56]  100.00,70.00,70.00,82.00,90.00,90.00,88.00
trainingData[57]  100.00,30.00,70.00,20.00,30.00,40.00,60.00
trainingData[58]  100.00,85.00,70.00,85.00,70.00,70.00,75.00
trainingData[59]  100.00,39.98,70.00,50.00,53.00,90.00,77.00
trainingData[60]  100.00,6.70,70.00,40.00,70.00,50.00,95.00
```

```
testingData[0]  65.00,97.00,70.00,45.00,100.00,100.00,40.00
testingData[1]  100.00,43.50,60.00,50.00,50.00,50.00,50.00
testingData[2]  92.50,65.00,65.00,53.00,50.00,100.00,75.00
testingData[3]  100.00,35.43,70.00,40.00,86.00,90.00,100.00
testingData[4]  100.00,13.33,30.00,24.00,89.00,50.00,75.00
testingData[5]  100.00,85.00,30.00,40.00,50.00,50.00,75.00
testingData[6]  100.00,45.00,70.00,34.00,70.00,50.00,53.00
testingData[7]  100.00,60.00,70.00,40.00,80.00,100.00,77.00
```

**Figure 3. Sample of Training and Testing datasets**

```
detector[0]   100.00,105.00,40.00,60.00,80.00,70.00,23.00
detector[1]   100.00,105.00,70.00,30.00,100.00,50.00,46.00
detector[2]   100.00,101.67,30.00,85.00,90.00,90.00,100.00
detector[3]   65.00,97.00,70.00,45.00,100.00,100.00,40.00
detector[4]   100.00,100.00,70.00,85.00,70.00,50.00,77.00
detector[5]   100.00,105.00,70.00,55.00,50.00,70.00,95.00
detector[6]   100.00,105.00,65.00,70.00,90.00,90.00,40.00
detector[7]   100.00,85.00,70.00,85.00,70.00,70.00,75.00
detector[8]   100.00,75.00,70.00,85.00,50.00,90.00,75.00
detector[9]   100.00,101.66,70.00,25.00,65.00,60.00,94.00
detector[10]  50.00,105.00,70.00,40.00,100.00,100.00,33.00
```

**Figure 4. List of Detectors**

The third phase is the model analysis phase, where in order to determine the accuracy of classification model, this study used 10 fold cross validation method for training and testing process. The accuracy of the proposed classification model using NS algorithm is represented by the average of accuracy for the dataset. The accuracy of the proposed model is based on the percentage of evaluation test data (untrained data) that are correctly classified.

## RESULT AND DISCUSSION

In this study, the accuracy of proposed classification model using NS algorithm was determined based on the 10 fold cross validation for the dataset. The number of new candidate that matches to the detectors will be considered as non-self that was used to calculate the accuracy of the model. As a result, the accuracy of the model is shown in Table 2, where the highest accuracy is 100% for 90:10 and 70:30 training ratio and the average of accuracy is 85.86% which is considered as a good classification model. In the classification process, especially in the model construction, the accuracy of the model should be higher or acceptable enough in order to produce a good model before it can be applied for classification or prediction on actual data.

**Table 2. The Accuracy of Classification Model**

| Training | Testing | Accuracy (%) |
|---|---|---|
| 90 | 10 | 100.00 |
| 80 | 20 | 66.67 |
| 70 | 30 | 100.00 |
| 60 | 40 | 80.00 |
| 50 | 50 | 77.78 |
| 40 | 60 | 90.00 |
| 30 | 70 | 83.33 |
| 20 | 80 | 91.67 |
| 10 | 90 | 83.33 |
| Average | | 85.86 |

As an example of application, the prototype system that applied NS algorithm to determine the right academician for promotion is shown Figure 5. This result shows the potential of NS algorithm for academic leadership classification for academic promotion exercise.

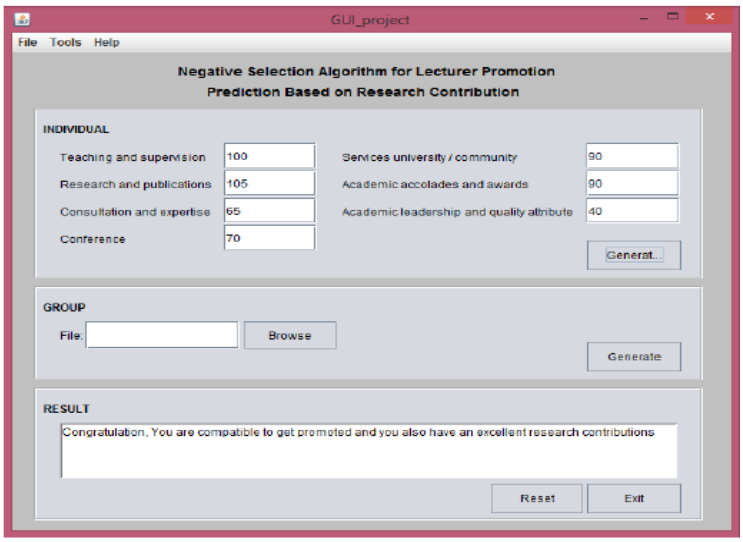



**Figure 5. Prototype System for Academic Leadership Classification**

## CONCLUSION AND FUTURE WORK

In this study, NS algorithm is proposed as a bio-inspired classification method for academic leadership classification. For future work, the proposed model can be proven by the comparative study using other artificial immune system such as Clonal Selection and Artificial Immune Network algorithms. It would give a direction on which algorithms can produce better result for academic leadership classification. As a conclusion, the ability to obtain new understanding of artificial immune network technique in human resource decision system will lead to the imperative contribution in human resource field.

## ACKNOWLEDGMENT

This work has been supported by Malaysian Ministry of Education (RAGS/2013/UiTM/ICT02/1).

## REFERENCES

Chang, J.-R., Cheng, C.-H., & Chen, L.-S. (2007). A fuzzy-based military officer performance appraisal system. *Applied Soft Computing, 7*, 936-945.

Forrest, S., Perelson, A. S., Allen, L., & Cherukuri., R. (1994). Self-nonself discrimination in a computer. *In Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, Los Alamos, CA,*.

Ismail, M., & Rasdi, R. M. (2008). Leadership In An Academic Career: Uncovering The Experience Of Women Professors. *Journal of the Commonwealth Council for Educational Administration & Management,, 3*(36).

Jantan, H., Hamdan, A. R., & Othman, Z. A. (2009, 25-27 February 2009). Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application. Paper presented at the *World Academy of Science, Engineering and Technology*, Penang, Malaysia.

Jantan, H., Hamdan, A. R., & Othman, Z. A. (2011). Data Mining Classification Techniques for Human Talent Forecasting. Paper presented at the *Knowledge-Oriented Applications in Data Mining*.

Ji, Z., & Dasgupta, D. (2007). Revisiting Negative Selection Algorithms. *Evolutionary Computation, 15*(2), 223-251.

Mohamed, S. A., Ammar, R. A., & Rajasekaran, S. (2012). Artificial Immune Systems: Models, Applications, and Challenges. Paper presented at the *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, New York, NY, USA.

Soam, S. S., Khan, F., Bhasker, B., & Mishra, B. N. (2011). *Identification of MHC Class II binders/ non-binders using Negative Selection Algorithm*.

Tarique, I., & Schuler, R. S. (2010). Global talent management: Literature review, integrative framework, and suggestions for further research. *Journal of World Business, 45*(2), 122–133. doi: 10.1016/j.jwb.2009.09.019

Timmis, J., Neal, M., & Hunt, J. (2000). An artificial immune system for data analysis. *Biosystems, 55*(1 - 3), 143–150.

Verhaegen, P. (2006). Academic Talent : Quo Vadis? Recruitment and Retention of Faculty in European Business Schools. *Journal of Management Development, 21*(9), 807-818.

Vincent-Lancrin, S. (2006). What is Changing in Academic Resarch?: Trends and Futures Scenarios. *Forthcoming in the European Journal of Education, 41*(2).

We, Y.-G., Zheng, D.-L., & Wang, Y. (2004). Research of A Negative Selection Algorithm and Its Application and Anomaly Detection. Paper presented at the T*hird International Conference on Machine Learning and Cybemetics*, Shanghai, China.