Identifying Online Credit Card Fraud using Artificial Immune Systems

A. Brabazon, J. Cahill, P. Keenan[1], D. Walsh,

UCD Business School,
University College Dublin,
Dublin 4, Ireland

***Abstract*—Significant payment flows now take place on-line, giving rise to a requirement for efficient and effective systems for the detection of credit card fraud. A particular aspect of this problem is that it is highly dynamic, as fraudsters continually adapt their strategies in response to the increasing sophistication of detection systems. Hence, system training by exposure to examples of previous examples of fraudulent transactions can lead to fraud detection systems which are susceptible to new patterns of fraudulent transactions. The nature of the problem suggests that Artificial Immune Systems (AIS) may have particular utility for inclusion in fraud detection systems as AIS can be constructed which can flag 'non standard' transactions without having seen examples of all possible such transactions during training of the algorithm. In this paper, we investigate the effectiveness of Artificial Immune Systems (AIS) for credit card fraud detection using a large dataset obtained from an on-line retailer. Three AIS algorithms were implemented and their performance was benchmarked against a logistic regression model. The results suggest that AIS algorithms have potential for inclusion in fraud detection systems but that further work is required to realize their full potential in this domain.**

I. INTRODUCTION

WebBiz (anonymized) conducts a growing online business in addition to their main street premises. WebBiz accepts payments for their online business through several channels such as credit cards, debit cards and bank transfers. This paper is concerned only with detecting fraud in credit card transactions made online. Although WebBiz employs good industry practice, credit card fraud remains a problem and occurs at a rate of about one fraudulent transaction in every million transactions. This paper describes the use of an Artificial Immune System approach in anomaly detection to attempt to reduce credit card fraud.

Credit card fraud is of significant economic importance, fraudulent card transactions in the U.S in 2005 were estimated to cost $790 million [1]. A UK survey of online businesses indicated that merchants expect to lose an average of 1.8% of their overall online revenue to payment fraud [2]. Credit card fraud can be broken down into two forms [3]; Inner card fraud requires collusion between merchants and cardholders and is not relevant here. External card fraud occurs when stolen, fake or counterfeit credit cards are used and this form of fraud is of interest here.
With the use of more secure "*Chip and PIN*" verification in most European countries, credit card fraudsters are increasingly targeting "*card not present*" transactions such as online shopping [4]. Data from the UK Payments industry shows a 118% increase in the value of phone, internet and mail order fraud (*card not present* fraud) between 2004 and 2008. From 2001 to 2008 *card not present* fraud losses in the UK rose by 243% and the total value of online shopping transactions increased by 524% [2]. Against this background of growing online transactions, which are less secure than over the counter transactions, there has been an increase in demand for credit card fraud detection. Fraud detection generally is an important area of application for artificial intelligence techniques [5] Anti fraud approaches for online credit card transactions have included the use of artificial intelligence, with new techniques being introduced in addition to older approaches, such as rule based systems [6]. There are commercial applications in this field, for instance the Falcon Fraud Manager software [7]. In addition, technical approaches such as enhanced encryption and passwords have been introduced in the credit card industry.

Artificial Immune Systems (AIS) are a recent branch of artificial intelligence based on the biological metaphor of the human immune system [8]. The immune system can distinguish between self and non-self, or more appropriately, between harmful non-self and everything else. This ability to recognize differences in patterns and to identify anomalies sparked the interest in adapting its processes to use in other domains, including the identification of anomalous credit card transactions.

The natural immune system is a highly complex system, comprised of an intricate network of specialized tissues, organs, cells and chemical molecules. The natural immune system can recognize, destroy, and remember an almost unlimited numbers of pathogens (foreign objects that enter the body, including viruses, bacteria, multi-cellular parasites, and fungi). To assist in protecting the organism, the immune system has the capability to distinguish between self and non-self. Notably, the system does not require exhaustive training with negative (non-self) examples to make these distinctions, but can identify items as non-self which it has never before encountered.

The negative selection algorithm was proposed in 1994 for anomaly detection [9]. The basis of the negative selection algorithm is the ability of the immune system to discriminate between self and non-self, or more broadly to

1. Peter..Keenan corresponding author *Peter.Keenan@ucd.ie*

distinguish between two system states, normal or abnormal. Forrest et al. [9] developed a binary-valued negative selection algorithm analogous to the negative selection or self-tolerogenesis process during T cell maturation in the thymus. Later this was extended to a real-valued representation. The process can be split into three stages, first the self cells need to be defined, and next a binary selection of cells is generated. These cells are randomly selected, the idea being that those who recognize the self samples contained in the training sample. The remaining detectors are used in the third stage of monitoring the occurrence of anomaly detection.

In implementing the algorithm, training data is usually normalized to (0,1) and a predetermined number of detectors are created at random positions in the data space. During the training process (akin to tolerogenesis) any detector that falls within a threshold distance r_{s} of any member of the set of self samples is discarded and replaced with another randomly generated detector. The replacement detector is also checked against the set of self samples. The process of detector generation iterates until the required number of valid detectors is generated. All of the resulting detectors are potentially useful detectors of non-self.

Once a population of detectors has been created they can be used to classify new data observations. To do this the new data vector is presented to the population of detectors and if it does not fall within a hypersphere of radius r_{s} of any detector, the data vector is deemed to be non-self. Otherwise, the new data vector is deemed to be self. A crucial point in the negative selection process is that the immune system does not require specific examples of non-self in creating its detectors. Potentially, the detectors can uncover any instance of non-self, even those never before encountered. This is an important attribute for fraud detection, as fraudsters continue to devise novel approaches to fraud which WebBiz may not have experienced before and the online nature of their business immediately exposes them to fraud innovation occurring anywhere in the world.

Many alterations of this model have been proposed; the major difference between the different models is the choice of a matching rule. This rule must determine the similarity between two patterns in order to classify self/non-self samples [10]. Other issues that weigh heavily with regards the success of a system are the number of detectors required, as well as the threshold set for the level of similarity. If this threshold value was to be too small it may not be possible to generate a suitable number of detectors from the available self. A balance needs to be found, setting the threshold high means the generated detectors become sensitive to any anomaly in the data patterns, so more detectors are necessary to achieve a desired reliability [10].

AIS approaches have been applied to several different areas. Different applications of information security have been examined through based on the workings of the immune system; these include host intrusion as well as network intrusion [11]. With host intrusion sequences of system calls were used as the detectors. Network intrusion has received a lot of research, the detectors here would relate to the IP address, whether it is the IP source address or the IP destination address. In the retail sector, a substantial research effort was undertaken using a computer intelligence fraud detection system [12]. This system consisted of a combination of negative selection as well as clonal selection. The results of this study were mixed. On the one hand, the AIS succeeded in highlighting different anomalies, however due to the large number of attributes or predictors for each transaction, the method of unsupervised learning proved troublesome. In 2008 Gadi et al [13] used data from a Brazilian Bank to study the effectiveness of using an Artificial Immune System to detect fraud. In their study, they used three different strategies and compared the results of using an Artificial Immune System, Artificial Neural Network, Decision Tree, Naive Bayes and Bayesian Nets with each of the strategies. The study focused more on reducing the cost of using each of the methods and producing the best set of parameters than on the success of each method with classification. There is no detail on which algorithm they used or which parameters were most successful.

In Tuo et al an Artificial Immune System for credit card fraud detection is suggested but not actually implemented [14]. It suggests integrating a Case based Reasoning approach with an Artificial Immune System. Brabazon, et al [15] study the use of Artificial Immune System in corporate failure prediction. In this study, both a canonical negative selection algorithm and a variable size detector algorithm were used and their performance compared. In this paper, the Artificial Immune System outperformed Linear Discriminant Analysis and, although it performed less well than the sophisticated GA/ANN approach, it has the advantage of not needing to be exposed to bad exemplars

II. Problem Data

This analytics uses data drawn from data provided by WebBiz which recorded 4 million transactions from 462279 unique customers with 5417 fraudulent transactions classified as fraudulent. Data was provided about the customer accounts, e.g. data of registration, and individual transactions e.g. date and amount of transaction (Table 1).

Data cleansing is a huge part of any project that involves raw data. It requires a huge amount of time and attention to ensure the quality of the overall data. They included having too much or too little data, missing data and noisy data or outliers. Initially the data was in two database tables and appropriate operations were needed to join these. Within the two files there were a number of transactions that were present in one but absent in the other i.e. missing data. These transactions could not be used since too many fields were absent. For this reason, they were eliminated. Since declined transactions have already failed a fraud prevention system, these too were removed from the data file.

TABLE I
PROBLEM DATA

Encrypted Customer ID	Customer identifier
Journal ID	Unique ID for the journal entry.
Date	The date and time of the transaction
Transaction type	The type of transaction carried out.
Reference key	Gives information about the type of transaction carried out. A separate table is provided which lists the different keys
User ID	Again, the most important aspect that this ID played was enabling the tracking of data. Very important when trying to normalize transactions with information split between the two files.
Amount	The transaction amount.
Description	Some additional information on the transaction
Balance	The customer's balance after the transaction
Payment ID	Again, the most important aspect that this ID played was enabling the tracking of data. Very important when trying to normalize transactions with information split between the two files.
Payment Sort	Identifies whether the transaction is a deposit or a withdrawal
Status	Y or N depending on whether the transaction was a success or declined
Chargeback Flag	A chargeback flag is needed to identify the known fraudulent cases. These transactions will be used for training purposes.
Scheme	The type of credit card used, e.g. MasterCard, Visa, etc.
Currency	The currency used.
Country of Bank	The country of the bank which was used to carry out the transaction is listed here
Secure 3D Flag	States whether the transaction was a 3D secure transaction or not
IP Address	The IP address from where the transaction took place.
Registration date	The date the customer registered to open their account.
Registration time	The time the customer registered to open their account.
Date of Birth	The customer's date of birth.

The most time consuming aspect of the data cleansing was dealing with noisy data. Different pieces of information such as country had been entered into the computer system manually. Hence, there were problems when it came to identifying the number of transactions in different countries, particularly the U.K, where there are several different representations i.e. G.B, England, U.K. etc. The IP address data also required a great degree of further preparation. As the four octet IP address e.g. 121.2.121.21 had to be matched with its location using the an IP to Country Database [16]. For ease of analysis, the date of birth was replaced by the age of the customer at the time they registered.

Before any normalization could be carried out, the information present needed to be standardized. Since WebBiz allow transactions in multiple currencies, these needed to be brought to a common currency.

There were three separate classes of variable in the final data, nominal, binary and ordinal or continuous data. Since all combinations of data were going to be examined, the normalized range needed to be standard and since the binary variables were present, the range of 0 to 1 was deemed most appropriate.

The continuous variables were normalized using an approach by Yu [17]

$$I = I_{min} + \frac{(I_{max} - I_{min})(D - D_{min})}{D_{max} - D_{min}}$$

I_{min} = *the minimal value of the input range, 0*.
I_{max} = *the maximal value of the input range, 1*.
D_{min} = *the minimal value of a given input range, varies for each variable, listed above.*
D_{max} = *the maximal value of a given input range, listed above*
D_{id} = *the figure due to be normalized*.

Since the input range is 0-1, this equation simplifies to

$$I = \frac{D - D_{min}}{D_{max} - D_{min}}$$

III. REGRESSION

We initially sought to apply conventional statistical techniques to the data, to provide a benchmark against which AIS techniques can be measured. Suitable benchmarks would include linear discriminant analysis (LDA) and logistic regression (LR).

Our dataset contains nominal, ordinal and binary data. For various reasons even the continuous variables do not have a close to normal distribution, for instance the age distribution is not normal as people under 18 generally do not have credit cards. Other data such as the date of first registration had a roughly uniform distribution (Fig 1). Consequently, linear regression is not an appropriate technique to use to investigate the data and we chose to use logistic regression. The logistic model calculates the probability of a certain outcome based on the values of the predictor variables.

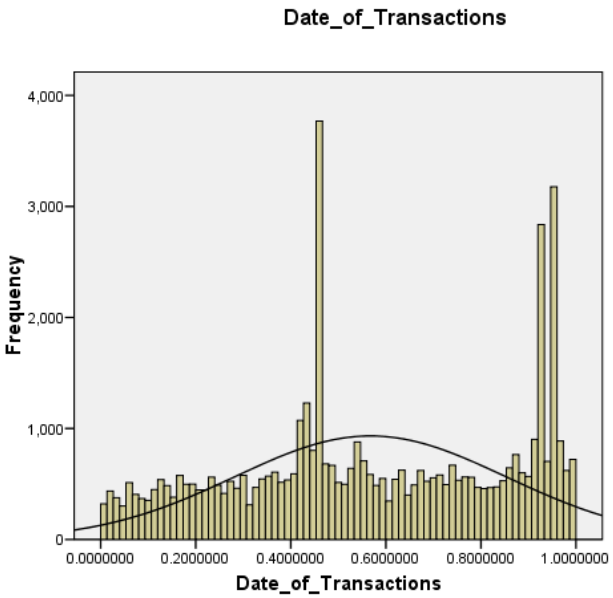

Figure 1: Histogram of Date of Transaction

Let the conditional probability that the outcome is present be denoted by

$$P(Y = 1|X) = \pi(x)$$

where *x* is the collection of predictor variables.

the logit of the multiple logistic regression model is given by

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

where β is estimated using maximum likelihood.

The maximum likelihood returns values for the unknown parameters which help to maximize the probability of obtaining the observed data set. The logistic regression model [18] can now be expressed

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

For the analysis in the study we used a randomly selected subset of 50,000 transactions. Of these, 49791 were non-fraudulent and 209 were fraudulent. Consequently the test data set has 0.418% transactions which were fraudulent and so has a much higher proportion of fraudulent transactions than the population

Using the logistic regression model 0.012% of transactions which were not fraudulent were misclassified. In addition, 85.167% of fraudulent transactions were misclassified. Overall using the logistic regression model 99.632% of transactions were accurately classified, compared to an accuracy of 99.582% for the naive method, suggesting that there are hidden relationships in the data and that an artificial immune system approach will yield useful results.

IV. Artificial Immune System approaches

The regression test results were then compared with three algorithms from the Artificial Immune Systems literature which we coded and tested on our data set. The three chosen algorithms were the Unmodified Negative Selection Algorithm, the Modified Negative Selection Algorithm and the Clonal Selection Algorithm. Before we consider these algorithms however, we need to decide on the appropriate distance The most commonly used distance algorithms in Artificial Immune Systems is the Euclidean distance algorithm [19]:

$$Euclidean\ distance(x,y) = \frac{1}{\sqrt{G}}\sqrt{\sum_{g=1}^{G}\left(\frac{(x_g - y_g)}{range_g}\right)^2}$$

However, this algorithm is not suitable for nominal variables. The fact that our data set contains nominal, ordinal and binary variables means that we need to investigate non-Euclidean distance measures to find the appropriate way to measure the distance between transactions in our dataset.

The concept that was most suited to our data set was the Value Distance Metric [20]. This metric is used to calculate the distance between nominal variables. It works using the probability that the value we are working with would be observed in our dataset. We will, however, use the unweighted version as presented in Wilson & Martinez.[21]

$$vdm_a(x,y) = \sum_{c=1}^{C}\left|\frac{N_{a,x,c}}{N_{a,x}} - \frac{N_{a,y,c}}{N_{a,y}}\right|^q \sum_{c=1}^{C}\left|P_{a,x,c} - P_{a,y,c}\right|^q$$

Where:

$N_{a,x}$ = the number of instances in the training set *T* that have value *x* for attribute *a*

$N_{a,x,c}$ = the number of instances in T that have value *x* for attribute a and output class *c*

C = the number of output classes

q = a constant, usually 1 or 2

$P_{a,x,c}$ = is the conditional probability that the output class is c given that attribute a has the value *x*

In order to apply a distance metric to our entire dataset, we use the Heterogeneous Value Distance Metrics drawn from Hanmaker & Bogess [19]:

$$HVDM(x,y)\frac{1}{\sqrt{G}}\sqrt{\sum_{g=1}^{G} hvdm\left(x_g, y_g\right)^2}$$

$$hvdm(x_g, y_g) = \begin{cases} \sqrt{vdm(x_g, y_g)} \\ \frac{|x_g - y_g|}{range_g} \end{cases}$$

Because our data includes both nominal and continuous data fields, we have to carefully consider how to "move" a detector away from itself. We considered using a very low mutation rate, generating a random number for each nominal data field and if the random number if below the mutation rate, randomly change it to one of the other possibilities for that field. We felt however that changing a nominal variable was not just moving a detector but changing it outright. For that reason we decided to adapt each of the continuous variables and leave the nominal variables as they were.

The decision was made to make the adaptation rate a random very small number and to randomly subtract or add to the variables. Again, when it came to moving the detector away from all other detectors when it was accepted, we decided that this was unnecessary in our version of the code. Because of the large number of possible states for some of the data fields we felt that there was sufficient variability in our detector set to ensure enough coverage.

V. Experimentation

In order to being testing our unmodified negative selection algorithm, it was necessary to first investigate the distances that would be returned by the HVDM distance metric.

After multiple iterations we found that the distances were all in the range [0, 0.2] with clustering around 0.1. This gave us a starting point for our radius. We decided to proceed as follows with our testing:

1. Vary the number of training samples
2. Vary the number of detectors
3. Vary the training radius
4. Vary the testing radius

We started with 200 detectors and setting the testing radius equal to the training radius, 0.055. The resulting detectors were tested on a sample of 1,000 transactions. Having tested various training samples sizes, the decision was made to proceed with the testing using 1,000 training samples.

We then tested using differing numbers of detectors. As the number of detectors was increased, the number of self transactions being classified as non-self was increasing but the detectors were still failing to identify almost any non-self transactions. At this point we decided to proceed with 200 detectors as that resulted in a low number of self being classified as non-self, as well as having a low run time.

By testing varying values for the radius, the best results were achieved with a radius of 0.07. As the radius is increased past this point, an unacceptably large number of self transactions are classified as non-self. At this point, the test radius is fixed at 0.07 and the testing radius is varied, where 0.6 was selected as the best value.

VI. Discussion of Results

The results of our analysis is shown in Table 2. The Unmodified Negative Selection Algorithm looks very promising if only the Accuracy result is examined. However, in order to gain a true understanding of how well this algorithm performed it is necessary to examine both figures that present the misclassification results. It turns out that while it classifies almost all of the self-transactions correctly, it misclassifies almost all of the non-self transactions. Because fraud is such a rare event, if one was to employ the scheme of classifying all transactions as non-fraudulent, one would achieve high levels of accuracy. This practice is obviously not a successful strategy for the identification of fraud however. For this reason, we have to conclude that the Unmodified Negative Selection Algorithm is not suitable for the detection of credit card fraud.

The Modified Negative Selection Algorithm offers a good tradeoff between classifying self correctly and classifying non-self correctly. As opposed to the Unmodified Negative Selection Algorithm which fails to classify any fraudulent transactions, the Modified Negative Selection Algorithm successfully identifies several of the fraudulent transactions.

TABLE 2
AIS RESULTS

Algorithm	Cut Number	% self categorized as non-self	% non-self categorized as self	Accuracy	Time taken (seconds)
Unmodified Negative Selection	1	0.4828%	96.55%	98.96%	5.9013
	2	0.945%	100%	98.48%	5.7366
	3	2.57%	100%	96.86%	5.5672
Modified Negative Selection	1	9.35%	96.55%	90.14%	142.186
	2	6.14%	89.66%	93.38%	143.653
	3	4.06%	96.55%	95.4%	138.766
Clonal Selection	1	19.73%	75.86%	79.94%	244.201
	2	33.8%	72.41%	65.98%	253.724
	3	39.29%	62%	60.72%	242.982

This result is very promising as it means that the Modified Negative Selection Algorithm has the ability to identify fraudulent transactions.

The results of the Clonal Selection Algorithm are very unpromising in relation to the accuracy metric used. It misidentifies a very high percentage of the self-transactions, and the nature of fraud is that there will always be many more normal transactions than fraudulent ones. This algorithm also takes the longest to run. However, the Clonal Selection Algorithm classifies a large percentage of the fraudulent transactions correctly, which is an extremely valuable trait in a fraud detection technique. An economic measure would have to take account of that the loss for a fraudulent transactions greatly exceeds the typical profit margin available for a successful non-fraudulent transaction. The large number of normal transactions likely to be misclassified makes this algorithm unsuited to fully automatic operation. However potentially fraudulent transactions could be subjected to further automatic or human processing to reduce the number of false negatives. In general, online customers are not willing to tolerate delay in credit card use online [4]. But there can be scope for review of customer accounts in the period before orders are fulfilled.

VII. CONCLUSIONS

The results suggest that AIS can be applied in this domain, but that care needs to be taken in the implementation of the algorithms in order to get workable results. Although the results obtained from the canonical Negative Selection Algorithm have high overall accuracy, the system misclassified too many fraudulent transactions to be operationalized. Several opportunities are indicated for future work. Design of an appropriate cost function which trades off the relative cost of Type I vs Type II errors is clearly important if a workable system is to be constructed. Another avenue is to investigate the utility of a hybrid, multistage detection system. Routine ‘fraud’ flags / rules could be applied to weed out the most obvious fraudulent transactions, leaving AIS to detect more subtle cases of fraud.

Future work might investigate the automatic generation of parameters using techniques such a genetic programming and genetic evolution [22].

REFERENCES

[1] S. Panigrahi, et al., "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning," Information Fusion, vol. 10, pp. 354-363, 2009.

[2] APACS. (2009, 05 Feb). APACS 2008 Fraud Loss Figures. Available: http://www.ukpayments.org.uk/media_centre/press_releases/-/page/685/

[3] A. Shen, et al., "Application of Classification Models on Credit Card Fraud Detection," 2007, pp. 1-4.

[4] S. Herbst-Murphy, "Maintaining a safe environment for payment cards: Examining evolving threats posed by fraud," Payment Cards Center Discussion Paper, 2009.

[5] [C. Phua, et al., "A comprehensive survey of data mining-based fraud detection research," Artificial Intelligence Review, 2005.

[6] K. J. Leonard, "The development of a rule based expert system model for fraud alert in consumer credit," European Journal of Operational Research, vol. 80, pp. 350-356, 1995.

[7] Falcon.(2010). (<http://www.fico.com/en/Products/DMApps/Pages/FICO-Falcon-Fraud-Manager.aspx>).

[8] L. N. De Castro and J. Timmis, Artificial immune systems : a new computational intelligence approach. London: Springer, 2002.

[9] S. Forrest, et al., "Self-nonself discrimination in a computer," in 1994 IEEE Symposium on Research in Security and Privacy, Los Alamitos, CA, USA, 1994, pp. 202–212.

[10] D. Dasgupta, et al., "Artificial immune system (AIS) research in the last five years," 2003, pp. 123–130.

[11] R. Overill, "Computational immunology and anomaly detection," Information Security Technical Report, vol. 12, pp. 188-191, 2007.

[12] J. Kim, et al., "Design of an artificial immune system as a novel anomaly detector for combating financial fraud in the retail sector," in CEC '03. The 2003 Congress on Evolutionary Computation, 2003. , 2003, pp. 405-412.

[13] M. Gadi, et al., "Credit Card Fraud Detection with Artificial Immune System," in Artificial Immune Systems, ed, 2008, pp. 119-131.

[14] T. Jianyong, et al., "Artificial immune system for fraud detection," in Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004, pp. 1407-1411 vol.2.

[15] A. Brabazon, et al., "Financial Classification using an Artificial Immune System," in Business Applications and Computational Intelligence, K. E. Voges and N. K. Pope, Eds., ed: Idea Group Publishing, 2006, pp. 389-406.

[16] IP. (Retrieved June 12, 2009). The IP to Country Database. Available: http://ip-to-country.webhosting.info/

[17] L. Yu, et al., "An integrated data preparation scheme for neural network data analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 18, pp. 217-230, 2006.

[18] D. W. Hosmer and S. Lemeshow, Applied logistic regression, 2nd ed. ed. New York ; Chichester: Wiley, 2000.

[19] J. S. Hamaker and L. Boggess, "Non-Euclidean distance measures in AIRS, an artificial immune classification system," in Evolutionary Computation, 2004. CEC2004. Congress on, 2004, pp. 1067-1073 Vol.1.

[20] C. Stanfill and D. Waltz, "Toward memory-based reasoning," Commun. ACM, vol. 29, pp. 1213-1228, 1986.

[21] D. R. Wilson and T. R. Martinez, "Improved Heterogeneous Distance Functions," Journal of Artifical Intelligence Research, vol. 6, pp. 1-34, 1997.

[22] H. Bernardino and H. Barbosa, "Grammar-Based Immune Programming for Symbolic Regression," in Artificial Immune Systems, ed, 2009, pp. 274-287.