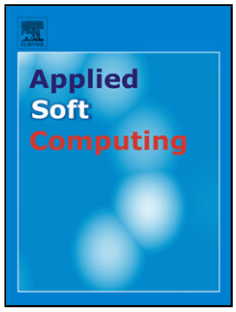
Accepted Manuscript



Title: An Artificial Immune Classifier for Credit Scoring Analysis

Authors: Shiow-Yun Chang, Tsung-Yuan Yeh

PII: S1568-4946(11)00426-1
DOI: doi:10.1016/j.asoc.2011.11.002
Reference: ASOC 1374

To appear in: *Applied Soft Computing*

Received date: 21-9-2009
Revised date: 9-8-2011
Accepted date: 1-11-2011

Please cite this article as: S.-Y. Chang, T.-Y. Yeh, An Artificial Immune Classifier for Credit Scoring Analysis, *Applied Soft Computing Journal* (2010), doi:10.1016/j.asoc.2011.11.002

An Artificial Immune Classifier for Credit Scoring Analysis

ABSTRACT

The primary concern of the rating policies for a banking industry is to develop a more objective, accurate and competitive scoring model to avoid losses from potential bad debt. This study proposes an artificial immune classifier based on the artificial immune network (named AINE-based classifier) to evaluate the applicants' credit scores. Two experimental credit datasets are used to show the accuracy rate of the artificial immune classifier. The ten-fold cross-validation method is applied to evaluate the performance of the classifier. The classifier is compared with other data mining techniques. Experimental results show that for the AINE-based classifier in credit scoring is more competitive than the SVM and hybrid SVM-based classifiers, except the BPN classifier. We further compare our classifier with other three AIS-based classifiers in the benchmark datasets, and show that the AINE-based classifier can rival the AIRS-based classifiers and outperforms the SAIS classifier when the number of attributes and classes increase. Our classifier can provide the credit card issuer with accurate and valuable information of credit scoring analyses to avoid making incorrect decisions that result in the loss of applicants' bad debt.

Keywords: Artificial immune network; Classifier; Credit scoring; Data mining; Classification

Shiow-Yun Chang, Tsung-Yuan Yeh

Department of Industrial and Information Management, National Cheng Kung University, Taiwan, ROC

1. Introduction

With the trend of internationalization and liberation of finance, many banking industries have expanded into the personal consumer finance market, especially in credit cards, which bring high profits via annual fees and revolving interest. Hence, the promotion of credit card businesses has become a very important strategy for generating revenue. To ensure the revenue from consumer finance, developing appropriate credit risk management and making good rating policies have both become important issues for the banking industry. The primary concerns of rating policies include how to avoid losses from potential bad debt, and the development of a more accurate and competitive scoring model. In the past, most credit card issuers used their experience or credit scoring system to assess customers' credit; hence, scoring is tallied by the subjective judgments of the issuer. The banking industry needs to reduce losses due to personal factors and labor costs, strengthen its credit risk management, and use information technology to assist or replace the credit issuers to construct an objective and rapid scoring system. Our study proposes a novel technique of evolutionary computation, which is extracted from the artificial intelligence field, namely the artificial immune system, to evaluate applicants' credits for improving the performance of credit scoring.

Credit scoring is based on a systematic analysis of the individual elements for the quality inspection of applicant credit. The first credit scoring technique, which is a simple parametric statistical method, is linear discriminant analysis [26]. Today, with the growth in the number of credit cards, the banking industry is developing more accurate credit scoring models. Recently, numerous classification concepts, principles and methods in different areas of data classification have been developed. The techniques used in credit scoring are genetic algorithm (GA) [6], artificial neural networks (ANNs) [6, 7, 33], discriminant analyses [7], logistic regressions [7], rough set theory [8], k-nearest neighbor models [13], support vector machine (SVM) [14], fuzzy theory [21, 23] and decision trees [25]. These techniques are all effective according to different datasets of the banking industries.

A large body of literature exists on assessing the implementation of the above techniques in improving the performance of credit scoring. However, within that literature, there is a surprising lack of application of the artificial immune system. The artificial immune system is a contemporary topic in artificial intelligence, and is a novel classification technique that simulates the ability of the natural immune system of the human body to detect foreign cells. The natural immune system is a very complex defense mechanism consisting of organs and many immune cells (i.e., mostly of lymphocyte cells), and it can prevent infectious agents such as bacterium and parasite from invading the body. The immune system has two essential types of lymphocytes, named bone-marrow-dependent lymphocytes (i.e., B-cells) and thymus-dependent lymphocytes (i.e., T-cells). These two types of cells are rather similar, but differ with relation to how they recognize the antigens and play their functional roles. The functions of the B-cells are cloning, mutating and producing matched antibodies to eliminate the incursive antigens from the body via the procedures of recognition and stimulation. The T-cells play a role in the discrimination between the "self" (innocuous) cells and the "nonself" (deleterious) cells, assist the B-cells in producing antibodies, and suppress the redundant stimulation of B-cells. Correspondingly, there are two types of immunity-based algorithms: imitating the behavior of B-cells or copying the reaction of T-cell antigens. Dasgupta et al. [4] surveyed the major works in artificial immune system (AIS) during the last few years. Their survey has revealed that recent studies are focused on four major AIS-based algorithms, namely "clonal selection algorithms", "immune networks theory", "negative selection algorithms" and "danger theory and dendrite cell algorithms".

The clonal selection algorithm, which was proposed by Burnet [2], was first developed and named CLONALG by De Castro & Von Zuben [5] and was initially applied to perform pattern recognition and multi-modal optimization task solving. The network theory of the immune system is another algorithm based on the clonal selection theory. Jerne [16] initiated a mathematical model of immune network theory that dynamically maintains the immune

memory via feedback mechanisms. Based on Jerne's mathematical model, Perelson [24] developed the theory of idiotypic networks that is a mathematical framework employing immunology. The immune network system is constructed by a set of B-cells, links between those B-cells, and cloning and mutating activities that are performed on B-cell objects. An immune response is elicited when a B-cell encounters an antigen, and the antibody then tries to bind itself with the antigen, so that the latter one can be neutralized. The immune network theory is mainly applied to data clustering, classification and an on-line fault diagnosis of industrial plant systems [3].

Different from the B-cells actions, the negative selection algorithm, which simulates T cells and was designed by Forrest et al. [10], is a change-detection method based on the computational model of self-nonself discrimination. It uses the ability of the immune system to detect unknown antigens while simultaneously not reacting to the "self" (innocuous) cells. The negative selection algorithm has various real-world applications, and has generally focused on the problems of anomaly detection and computer security, such as network intrusion detection, virus detection and operating system monitoring [3]. Concurrently, Matzinger [22] introduced the danger theory, which is based on the co-stimulated model of allogeneic interactions. The main difference from the classical immune algorithms is that danger theory does not respond to all foreign cells, but only to those that are dangerous to the body. The danger theory was applied to anomaly detection, especially for the danger signals that conduct automatic measurements such as too low or too high memory usage [3].

The common characteristic of the above various immunity-based algorithms is a naturally strong ability of antigen recognition and antibody evolution derived from the human body. The numerical results of former research showed that the artificial immune algorithms have been applied and developed in various fields, for instance, anomaly detection, computer security and virus detection, data classification and clustering, fault diagnosis, pattern recognition, scheduling and web mining [3, 9, 12]. For data classification and clustering, an

artificial immune algorithm, clustering analysis and self-origination maps neural network are used to classify Fisher Iris dataset. The numerical results confirmed that an artificial immune algorithm is the most effective technique [33]. Subsequently, Timmis et al. [27] first proposed the artificial immune system as an unsupervised clustering tool, but also stressed its use as an exploratory data analysis and visualization technique for a dataset with four-dimension. They suggested that the artificial immune system would involve the application of the algorithm to more complex problems of higher dimensional datasets. Leung et al. [20] compared the classification performance of some classifiers (e.g., ANNs, SVM, etc.) against an artificial intelligence technique based on the natural immune system, named the simple artificial immune system (SAIS), through three credit datasets. They showed that the simple artificial immune system is a competitive classifier.

Since the artificial immune system has merits in recognition and evolution, the initial idea of this study is based on the principles and abilities of an immune system that can identify the innocuous cells (i.e., applicant with good credit) and the deleterious cells (i.e., applicant with bad credit). Our study is mainly concerned with the data mining technique and focuses on the classification of credit applicants, and develops an artificial immune classifier in credit scoring. The classifier is tested by using ten-fold cross-validation with two real world credit datasets of the banking industries, and is compared with the techniques of ANN, decision trees, Na ve-Bayes, SVM, hybrid SVM-based and SAIS, and with other AIS-based classifiers for the benchmark datasets. Finally, we verify that the proposed classifier indeed can successfully classify an applicant as approved for good credit or rejected for bad debt, and achieve better performance than traditional statistical methods. Therefore, the classifier is a suitable and competitive classifier in credit scoring.

The remainder of this paper is organized as follows. In section 2, the artificial immune classifier is described. Section 3 presents two real world credit datasets of the banking industries that are used to evaluate the classification performance of the classifier. After that,

the accuracy rate of the classifier is compared with others in Section 4. The final section offers some concluding remarks.

2. Artificial immune classifier

Our study modifies the original model of Timmis et al. [27] based on the artificial immune network to develop an artificial immune classifier called "AINE-based classifier" for credit scoring. When a receptor of B-cell finds its antigen, namely by having close connection between these two cells, the immunoreaction impels this B-cell to split at this moment to produce the cells with the same recombinant genes, which is the concept of cloning and mutating behavior. Whether the B-cells could produce good immunoreactions with the antigen is determined by the value of the stimulation level. The B-cells, of which their stimulation levels exceed the threshold value of simulation, are transformed into blast cells that divide and produce clones and mutations of the B-cells until the maximum stimulation level of the B-cells exceeds the evolution termination value of the network system. The behavior of the clones and mutations for the B-cells can strengthen immune memory capacity and recognize the slightly different antigen.

In our model, we consider the evolution termination value of network system in order to terminate cloning and mutating to generate the matched data items. This value is regard as a threshold of the system to ensure that the system does not repeatedly proceed to clone and mutate when the matched data items have been found, and thereby avoids generating a large cloned data items (e.g., cloned B-cells). Note that in this study the weakest 5-10% of the B-cells are removed from system at each iteration in order to fit the characteristic of the AIS-based classifier [27].

The AINE-based classifier uses the following notation.

A the constant scalar used to control connectivity, $0 \le A \le 1$

m the mutation value, $0 \le m \le 1$

k a scalar constant of clone

NET the evolution termination value of the network system

N the number of the B-cells

B the initial number of the B-cells

SL the stimulation level of the B-cell

ps the stimulation between the B-cell and the antigen

SS the stimulation between the B-cell and its neighboring B-cells

sf suppression factor, the suppression between the B-cell and its neighboring B-cells

dis_a the Euclidean distance between B-cell i and antigen a, $0 \le dis_a \le 1$

dis_x the Euclidean distance between B-cell i and B-cell x, $0 \le dis_x \le 1$

aff_j the connecting intensity between two cell objects associated with the j^{th} link

NST the threshold value of stimulation for the network system

$SL_{\max}$ the maximum stimulation level of the B-cell

e the mutated clonal number of the B-cell

In the immunology, the calculation of the affinity and the stimulation level can determine whether the B-cell could effectively protect against the incursive antigen. Affinity represents how well the two cell objects are matched and bound. The surface of B-cell contains the antibody for that B-cell. When an antibody for an initial B-cell tightly binds (i.e., strong affinity) to an incursive antigen, the initial B-cell becomes stimulated and then produces the matched antibody to eliminate the incursive antigen. On the other hand, stimulation level (SL) represents that how well a B-cell matches with the antigen and its affinity to the other B-cells in the network system. The stimulation level is according to the stimulation between the antigen and the B-cell (ps), the stimulation between the B-cell and its neighboring B-cells (SS), and the suppression between the B-cell and its neighboring

B-cells (sf) in the immune system. In addition, a higher stimulation level of the B-cells results in the system to produce a large number of clones, and the behavior of clone and mutation can improve the matching ability for the antigen. When the antibody and the incursive antigen are matched and bound and the value of stimulation easily exceeds the certain threshold value, the system proceeds to clone and mutate a number of B-cells that are similar to the features of the original B-cell, which increases the likelihood of the antibody to match with the antigen.

The sketch of the AINE-based classifier is shown in Fig. 1. To imitate the immune system, the data item of the system can be regarded as the antibody for the initial B-cell, and the data item of the applicant can be regarded as the incursive antigen. When the antigen (i.e., data item of the applicant) are bound and matched with the initial B-cell (i.e., data item of the system), the system produces the matched antibodies (i.e., classification results). Note that the data values of the dataset are normalized before implementation.

The algorithm of AINE-based classifier in credit scoring is shown in Algorithm. To begin with, we set the fixed parameters, A, m, k, NET and B, and then the network system randomly generates the initial data items (i.e., system antibodies, B-cells), Then, we input the data items of an applicant into the system to match with the initial data items. After that, the system calculates the stimulation levels of all the initial data items and the threshold value of stimulation of the network system (NST) before classifying the data items of the applicant. To find the matched data item, if the value of maximum stimulation level ($SL_{\max}$) is larger than the controllable pre-set evolution termination value of network system (NET), then the system stops cloning and mutating immediately, and generates the matched data item. Otherwise, the system clones and mutates all the B-cells for which their values of stimulation levels exceeds the threshold value of stimulation, and then combines with the original data items into a brand-new set. Finally, system repeats the above procedure until all the matched data items of the applicants are found.

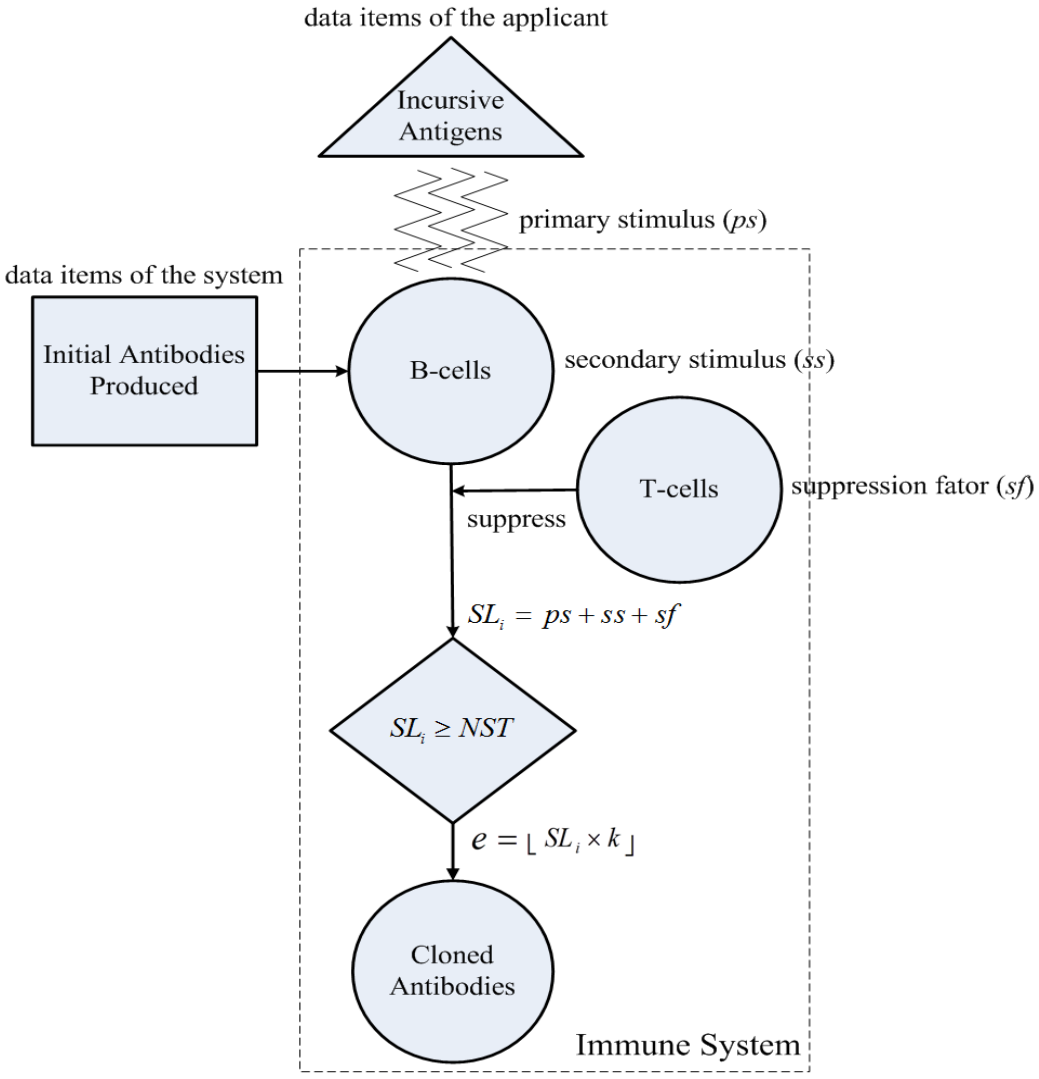


Fig.1. The sketch of the AINE-based classifier.

Algorithm (AINE-based Classifier)

Initialization: Set A, m, k, NET and B, and the network system randomly generate B initial B-cells and let $N = B$.

Step 1: Input the data item of an applicant.

Step 2: The calculation of SL_i for each B-cell i

$$SL_i = ps + ss + sf$$

$$= (1 - dis_a) + \sum_{x \neq i} (1 - dis_x) - \sum_{x \neq i} (dis_x)$$

Step 3: The calculation of NST

$$NST = \frac{A \sum_{j=1}^{N} (1 - aff_j)}{N}$$

Let $SL_{\max} = \max_{1 \leq i \leq N} \{SL_i\}$

Step 4: IF $SL_{\max} \geq NET$

Output the matched data item of the applicant.
GoTo Step 5
Else
IF $SL_i \geq NST$ for each B-cell i
$e = \lfloor SL_i \times k \rfloor$
$N = N + e$
GoTo Step 2
End
End
Step 5: GoTo Step 1 until all the matched data items of the applicants are found.

3. Experimental results and discussions

The University of California at Irvine (UCI) maintains a machine-learning repository of the datasets [1] for the development and testing of classification algorithms. Our study uses the two UCI datasets concerning credit application to test the AINE-based classifier. To begin with, the classifier sets the relative parameters, and then the system evaluates the classification performance of two datasets by a two-class confusion matrix.

The experimental datasets used in this study are Australian credit approval and German credit datasets. The Australian credit dataset has 690 instances, with 6 numerical attributes, 8 categorical attributes and 2 classes (rejected and accepted), 383 of which represent individuals whose credit card applications are rejected (55% of total instances) and the remaining 307 individuals are approved for their credit card applications. On the other hand, the German credit dataset has 1000 instances, with 7 numerical attributes, 13 categorical attributes and 2 classes (rejected and accepted), 300 of which represent individuals whose credit card applications are rejected (30% of total instances) and the remaining 700 individuals are approved for their credit card applications. Note that for the two datasets we do not know the meanings of attributes or which attribute is important, because all the input attributes' names and values have been changed into symbols.

3.1. Setting parameters of the AINE-based classifier

Choosing the most suitable parameters combination is the first important process for the classifier, because it affects the computation time of the program, average-RMSE, network size, and the connectivity ability between the incursive antigen and the matched antibody [27, 28]. Timmis et al. [28] indicated that the network size and connectivity ability are restricted by the constant scalar used to control connectivity (A) and the mutation value (m), and significantly affects the number of the B-cells and the process time of evolution (i.e., computation time). To improve the classification performance, the referral literature, experiential judgment and trade-off decision between the computation time and average-RMSE are the critical factors when choosing the most suitable parameters combination. Hence, some experimental results [28] showed that the recommended values of the constant scalar used to control connectivity and the mutation are 0.5 and 0.1, respectively, however, in our model they are set to 0.8 and 0.05, respectively, while the initial number of the B-cells (B) is then set to 500.

The numerical results (see Table 1) show that the scalar constant of clone (k) and the evolution termination value of the network system (NET) are both positively related to the level of evolution. Our study considers both the computation time and average-RMSE at the same time, hence the scalar constants of clone for Australian credit approval and German credit dataset are then set to 3, and the evolution termination values for Australian credit approval and the German credit dataset are set to 0.6 and 0.7, respectively. The combination of these parameter values can make a stronger connection between the incursive antigens and the matched antibodies with acceptable computation time. For the Australian credit approval, the computation time is 366.9940 seconds, while the average-RMSE is 0.1348. For the German credit dataset, the computing time is 2896.4 seconds, while the average-RMSE is 0.2476. Here the computation time is based on a PC with an AMD Athlon XP CPU running at 1.15GHz with 256MB RAM.

From Table 1, we also find that the process time of evolution (i.e., computation time) increases with a scalar constant of clone, and with the evolution termination value of the network system. In addition, the average-RMSE decreases as the evolution termination value increases regardless of whether the scalar constant of clone is small or large. When the evolution termination value is small, the evolution process will roughly get up to the evolution termination value that resulted in weak connection between the incursive antigen and the matched antibody. On the contrary, the incursive antigen is strongly connected with the matched antibody when the evolution termination value is large, which resulted in the system repeatedly proceeds to evolve while the incursive antigen and the matched antibody have conjugated continuously, and thus average-RMSE will reduce relatively.

Table 1 The computation time and average-RMSE of the credit datasets.

Clone constant (k)	Evolution termination value (NET)	Australian credit approval		German credit dataset	
		Average-RMSE	Computation time (s)	Average-RMSE	Computation time (s)
	0.4	0.1832	172.2270	0.2970	437.9090
	0.5	0.1615	297.3280	0.2733	910.2300
3	**0.6**	**0.1348**	**366.9940**	0.2629	2001.0000
	0.7	0.1142	1093.2000	**0.2476**	**2896.4000**
	0.4	0.1827	187.9400	0.2962	453.1890
4	0.5	0.1608	308.5770	0.2731	930.5930
	0.6	0.1348	682.4450	0.2613	2319.5000
	0.7	0.1142	2990.5000	0.2477	3464.4000
	0.4	0.1821	143.8870	0.2705	507.0480
5	0.5	0.1612	310.3360	0.2792	2736.4000
	0.6	0.1338	1093.2000	0.2613	3993.1000
	0.7	0.1141	4175.8000	0.2477	6091.3000

3.2. The classification results of the AINE-based classifier

Our study uses the principal component analysis (PCA) to extract the attribute features. It is generally performed on data attribute reduction and classification, and provides a useful

comparison of the numerical results from the classifier [11, 27]. We apply PCA to the classification results (i.e., matched antibodies), and then make the confusion matrix [11]. A simple two-class confusion matrix is used to analyze the classification performance of the AINE-based classifier. The three measurements of a confusion matrix are the accuracy rate, sensitivity and specificity. The accuracy rate is the percentage of rejected and approval applicants who are correctly classified. The sensitivity is the percentage of applicants with bad credit who are correctly classified as having bad credit. The specificity is the percentage of applicants with good credit who are correctly identified as having good credit.

According to the matched antibodies of the Australian credit approval, the total instances show that a phenomenon of two well-separated clusters of approximately equal size. Here Cluster I and Cluster II represent most of rejected and accepted instances, respectively. The numerical results show that 327 matched antibodies of the rejected instances are classified into Cluster I, and 56 rejected instances are classified into Cluster II. Similarly, 285 matched antibodies of the accepted instances are classified into Cluster II and 22 accepted instances are classified into Cluster I, as shown in Table 2. In addition, the results of the German credit dataset are similar to the Australian credit approval, as shown in Table 3.

Table 2 The confusion matrix of the Australian credit approval.

Actual condition	Cluster I (349 instances rejected)	Cluster II (341 instances accepted)
Reject (383 instances)	327	56
Accept (307 instances)	22	285
Accuracy rate = 88.70%	Sensitivity = 85.38%	Specificity = 92.83%

Table 3 The confusion matrix of the German credit dataset.

Actual condition	Cluster I (369 instances rejected)	Cluster II (631 instances accepted)
Reject (300 instances)	232	68
Accept (700 instances)	137	563
Accuracy rate = 79.50%	Sensitivity = 77.33%	Specificity = 80.43%

Analyzing the above numerical results, the classification performance of our classifier for the Australian credit approval is better than that for the German credit dataset. Why do the two datasets have large differences in accuracy rates when used on the same classifier? The primary reason is the data of the Australian credit approval extensively decentralized in the network system. The extensive and decentralized data represents strong affinity, and the system thus procures the small threshold value of stimulation, which results in the value of stimulation level easily exceeds the threshold value of stimulation, hence the system repeatedly proceeds to clone and mutate until the system evolution terminated. The behavior of repeating clones and mutations results in the system to produce the extensive decentralized B-cells, which increases the likelihood for the antibody (i.e., data items of the system) to match with the incursive antigen (i.e., data items of the applicant) in the system. In other words, the incursive antigen thus strongly connects with the matched antibody, and consequently average-RMSE decreases and the accuracy rate is relatively higher.

Another point to be noted about the accuracy rate is the extreme values of the data. Unlike German credit dataset, there are many extreme values in the Australian credit approval. The system repeatedly proceeds to evolve and produces the B-cells according to the resemblance features of the extreme values, which leads to the increase of computation time and slightly affects the entire accuracy rate due to the erroneous classification. However, in order to keep the original dataset for its completeness, and in case the extreme values do not serious affect the classification results, we suggest that the extreme values of the attributes are worthy of being remained in the dataset when using this classifier.

4. Performance evaluation

In this section, we use a ten-fold cross-validation technique to examine the performance evaluation of the classifier, and apply stepwise regression and the Wilcoxon signed-rank test to see if there exists any difference after attribute selection. Finally, our classifier is compared

with the accuracy rate of other classifiers.

4.1. Performance evaluation of the classifier

Our study applies ten-fold cross-validation to partition each dataset into ten disjoint subsets, namely training-testing trial sets and the correctness of classifier is computed as the average accuracy realized from the training-testing trials. In general, even if computation power allows using more folds, a ten-fold cross-validation technique is recommended for the classifier to examine the accuracy due to its relatively low bias and variance [17].

After applying stepwise regression, there are 7 (categorical: 4; numerical: 3) and 10 (categorical: 7; numerical: 3) attributes selected from the Australian credit approval and the German credit approval dataset, respectively. For the Australian credit approval, the average accuracy rate of data classification is 85.36%, and increases to 86.38% after attribute selection. As regards the German credit dataset, the average accuracy rate of the data classification is 77.10%, and increases to 77.90% after attribute selection. Both the p-values of the Wilcoxon signed-rank test for the two datasets are larger than the significant level ($\alpha = 5\%$), as shown in Table 4. These results conclude that the datasets do not provide sufficient evidence to establish that there exists significant differences in the accuracy rates after attribute selection when used on the artificial immune classifier with the same parameters. These numerical results show that some input attributes are highly correlated, hence these attributes can be eliminated.

The analysis results suggest that the stepwise regression is good for attributes selection, and the AINE-based classifier in credit scoring after attribute selection still has a good classification performance. Therefore, for the two datasets our study indicates that both the number of the eliminated attributes and the data types do not affect the average accuracy rate of data classification, but decrease the computation time. Clearly, the above findings show that the AINE-based classifier indeed can be used to classify the applicants of credit scoring

and has a good classification performance.

Table 4 The numerical results of performance evaluation.

	Australian credit approval		German credit dataset	
	Original data	Attribute selection	Original data	Attribute selection
Average-RMSE	0.1348	0.1299	0.2476	0.2062
Average accuracy rate (%)	85.36	86.38	77.10	77.90
Computation time (s)	366.9940	122.7071	2896.4000	749.2950
Wilcoxon signed-rank test	p-value = 0.237		p-value = 0.067	

4.2. Comparison of the accuracy rates with other AIS-based classifiers

Next subsections, some recent accuracy rates of other three AIS-based classifiers for the benchmark datasets and that of other well-used classifiers for the credit datasets of the banking industries are compared with our classifier.

4.2.1. Benchmark datasets

The AINE-based classifier is compared with the accuracy rates of other three AIS-based classifiers for three well-known benchmark datasets, namely "Fisher Iris", "Johns Hopkins University Ionosphere" and "Pima Indians Diabetes" [1], which are obtained from the studies of Leung et al. [20] and Watkins et al. [32]. Each information of datasets is as follows: The Fisher Iris dataset has 150 instances with 4 numerical attributes and 3 classes (Setosa, Versicolour and Virginica). The Johns Hopkins University Ionosphere dataset has 351 instances with 34 numerical attributes and 2 classes (good and bad). The Pima Indians Diabetes dataset has 768 instances with 8 numerical attributes and 2 classes (positive and negative).

For the Fisher Iris and the Johns Hopkins University Ionosphere datasets, their most suitable initial constant scalar used to control connectivity and the mutation value are set to

0.5 and 0.1, respectively, while the initial number of the B-cells is set to 100. For the Fisher Iris dataset the scalar constant of clone and the evolution termination value of the network system are set to 5 and 0.7, respectively, and for the Johns Hopkins University Ionosphere dataset they are set to 4 and 0.8, respectively. Hence, the computation time and the average-RMSE for the Fisher Iris dataset are 50.2930 seconds and 0.0698, respectively, and those for the Johns Hopkins University Ionosphere dataset are 905.4300 seconds and 0.1695, respectively. In addition, for the Pima Indians Diabetes dataset the most suitable initial constant scalar used to control connectivity and the mutation value are set to 0.7 and 0.1, respectively, while the initial number of the B-cells is set to 300. The scalar constant of clone and the evolution termination value of the network system are set to 3 and 0.8, respectively. Hence, the computation time and the average-RMSE for Fisher Iris dataset are 4018.7000 seconds and 0.3015, respectively. Here the computation time is based on a PC with an AMD Athlon XP CPU running at 1.15GHz with 384MB RAM.

The numerical results are summarized in Table 5. For each benchmark dataset, the computation time is positively related to both the scalar constant of clone and the evolution termination value of the network system. In short, in Tables 1 and 5 we confirm that the process time of evolution increases with both a scalar constant of clone and the evolution termination value. The comparative results of all classifiers are summarized in Table 6. Note that both AIRS I and AIRS II (artificial immune recognition system) [31] are based on the concept of the artificial recognition balls (ARBs) and the principle of the resource limitation [29]. The numerical results show that the accuracy rates of our classifier for the Fisher Iris, Johns Hopkins University Ionosphere and Pima Indians Diabetes datasets are 96.0%, 89.2% and 74.9%, respectively. For both the Fisher Iris and Pima Indians Diabetes datasets the AINE-based classifier is less accurate than the SAIS classifier, however, for the Johns Hopkins University Ionosphere dataset it performs better than the SAIS classifier.

Table 5 The computation time and average-RMSE of the benchmark datasets.

Clone constant (k)	Evolution termination value (NET)	Fisher Iris		Johns Hopkins University Ionosphere		Pima Indians Diabetes	
		Average-RMSE	Computation time (s)	Average-RMSE	Computation time (s)	Average-RMSE	Computation time (s)
3	0.6	0.0732	32.8910	0.1973	121.9850	0.3175	777.9050
	0.7	0.0710	44.3160	0.1928	175.0460	0.3074	1145.9000
	0.8	0.0662	101.3880	0.1716	613.9010	**0.3015**	**4018.7000**
4	0.6	0.0732	35.9760	0.1952	132.4470	0.3099	1065.7000
	0.7	0.0709	47.1200	0.1904	203.1560	0.3016	2522.3000
	0.8	0.0577	182.5510	**0.1695**	**905.4300**	0.2994	6359.2000
5	0.6	0.0732	40.9050	0.1944	149.4520	0.3145	1904.3000
	0.7	**0.0698**	**50.2930**	0.1803	215.9840	0.3007	3005.9000
	0.8	0.0576	347.5000	0.1695	1049.4000	0.2993	7896.4000

Table 6 The accuracy rates of benchmark datasets for the AIS-based classifiers.

	Accuracy rates (%)		
Classifier	Fisher Iris	Johns Hopkins University Ionosphere	Pima Indians Diabetes
AINE-based (this study)	96.0	89.2	74.9
AIRS I	96.7	94.9	74.1
AIRS II	96.0	95.6	74.2
SAIS	97.3	87.5	77.4

According to the study of Leung et al. [20], for the Johns Hopkins University Ionosphere dataset the primary reason to the lower accuracy rate of the SAIS classifier is that the number of attributes and classes of the dataset affects the classification performance. However, we find that both the AINE-based and SAIS classifiers have this common characteristic, in which their accuracy rates tend to decrease when the number of attributes and classes of the dataset increase, however, our classifier is still more accurate than the SAIS one. On the other hand, the accuracy rate of the AINE-based classifier for the Johns Hopkins University Ionosphere dataset is less accurate than that of the AIRS-based classifiers, but it outperforms them for the

Pima Indians Diabetes dataset. For the Fisher Iris dataset the AINE-based and AIRS-based classifiers have no significant difference in the accuracy rate. Therefore, our study shows that the AINE-based classifier can rival the AIRS-based classifiers.

Analyzing the numerical results, we find that the classification performance of our classifier for the Pima Indians Diabetes dataset is worse than that for other two datasets. The primary reason is similar to the reason for the German credit dataset. The Pima Indians Diabetes This dataset has the extensive and decentralized data in the network system. The extensive and decentralized data represents weak affinity, and the system thus procures the large threshold value of stimulation, hence the system is then difficult to clone and mutate. In other words, the incursive antigen weakly connects with the matched antibody, and consequently average-RMSE increases and the accuracy rate is relatively lower.

4.2.2. Australian credit approval and German credit dataset

The AINE-based classifier is compared with SVM [14], hybrid SVM-based [15], backpropagation neural network (BPN) [18], Na ve-Bayes [19], SAIS [20], C4.5 [25] and the original model of Timmis et al. [27] by application of the same credit datasets.

For the model of Timmis et al., the most suitable initial constant scalar used to control connectivity and the mutation value for the two datasets are set to 0.5 and 0.1, respectively, while the initial number of the B-cells is then set to 500. The trade-off decision between the computation time and average-RMSE is that the scalar constants of clone for both the datasets are then set to 3 (see Table 7). For the Australian credit approval, the computation time is 2370.5000 seconds, while the average-RMSE is 0.1817. For the German credit dataset, the computing time is 4377.1000 seconds, while the average-RMSE is 0.3144. Here the computation time is based on the same PC conditions as the benchmark datasets.

Table 7 The computation time and average-RMSE of the model of Timmis et al.

	Australian credit approval		German credit dataset	
Clone constant (k)	Average-RMSE	Computation time (s)	Average-RMSE	Computation time (s)
2	0.1820	1293.0000	0.3150	2805.4000
3	**0.1817**	**2370.5000**	**0.3144**	**4377.1000**
4	0.1812	5094.7000	0.3139	7858.3000
5	0.1812	7977.3000	0.3139	9713.0000

The comparative results of all the classifiers are summarized in Table 8. The results show that both the accuracy rates of two datasets for the AINE-based classifier ranked fourth of all classifiers, however, for non-hybrid classifiers both the accurate rates are ranked at the second position, however, the AINE-based classifier is only less accurate than the BPN classifier. Comparing the three AIS-based classifiers, we find that for the Australian credit approval the AINE-based classifier can rival the model of Timmis et al. and the SAIS classifier, and for the German credit dataset it is the most accurate of all AIS-based classifiers.

For hybrid classifiers both the accurate rates of two datasets for the AINE-based classifier are less than the SVM-GA classifier, however, the accuracy rate of the AINE-based classifier is higher than the SVM-Grid-F-Score classifier and the SVM-Grid classifier for the Australian credit approval and the German credit dataset, respectively. Briefly, the AINE-based classifier performs better than the non-hybrid classifiers in substance, except the BPN classifier, and can rival both the SVM-Grid and SVM-Grid-F-score classifiers. The results also show that the AINE-based classifier is more accurate than the SVM classifier, but the hybrid SVM-based classifier is competing with the AINE-based classifier, however, the hybrid model generally requires more computation time. The experimental results [30] showed that the hybrid model could provide higher prediction accuracy with more computation time. Based on the same computer hardware (e.g., CPU/RAM), the computation

time of our classifier is possibly better than other hybrid classifiers, therefore, we suggest that the trade-off decision between the computation time and accurate rate may profitable in improving the classification performance when considering a hybrid model.

Table 8 The accuracy rates of various credit scoring techniques.

	Accuracy rates (%)	
Classifier	Australian credit approval	German credit dataset
AINE-based (this study)	85.36	77.10
BPN	86.83	77.80
C4.5	82.50	72.40
Na ve-Bayes	84.90	74.70
SAIS	85.20	75.40
SVM	84.70	76.00
Model of Timmis et al.	85.20	72.40
hybrid SVM-based model		
SVM-GA	86.90	77.90
SVM-Grid	85.51	76.00
SVM-Grid-F-score	84.20	77.50

From the past studies, the good initial values for a heuristic method or soft computing technique is a critical procedure in order to raise the evolution effect and then improve the accuracy rates. In addition, the ongoing work using simulated annealing and immune principles applied to the problem of finding good initialization values for neural networks was presented [8], and also, an immune system was applied to genetic algorithms for searching an optimum solution [5]. In terms of the above applications, our study suggests that the AINE-based classifier can be combined with other techniques, such as neural networks or genetic algorithms, to improve the classification performance of the original classifier.

5. Conclusions

This study proposes an AINE-based classifier based on the artificial immune network for credit scoring. Two UCI experimental credit datasets of the banking industries show that the

AINE-based classifier outperforms the SAIS and SVM the classifiers and the model of Timmis et al. Hence, the AINE-based classifier is indeed a suitable and competitive classifier in credit scoring. We further do some experiments in the three well-used benchmark datasets for other three AIS-based classifiers (i.e., AIRS I/II and SAIS), and the results show that our classifier can rival the AIRS-based classifiers. We find that both the accuracy rates of the AINE-based and SAIS classifiers tend to decreases when the number of attributes and classes increase, however, the AINE-based classifier is more accurate than the SAIS classifier in such a dataset.

In this study, we confirm that the process time of evolution (i.e., computation time) increases with both a scalar constant of clone and the evolution termination value, and the extensive and centralized data in the dataset significantly affect the classification performance of all the AIS-based classifiers. In addition, the numerical results indicate that the attribute selection do not affect the accuracy rates of data classification, but affects the computation time when applied to the two datasets.

The results from this study suggest three main aspects that may in further improving the classification performance of the original classifier. The first aspect is determining the trade-off analysis between the computation time and classifier type (i.e., a non-hybrid or hybrid model) when using a complex hybrid mode. Second, in this study the eliminated attributes do not affect the accuracy rate, however, based on some limitations of data types for various techniques, the questionnaires design and collection related the information of the applicants ought to consider the follow-up data analyses. For example, the categorical data, which is classified as nominal or ordinal data, resulted in few analytic methods to analyze them, except frequency concept. Finally, other immune-based algorithms can be applied in credit scoring, or the hybridized AINE-based classifier can be used to improve the classification performance of the original classifier.

Acknowledgements

The authors would like to thank the anonymous referees for their careful reading and valuable comments which helped to improve this study.

References

[1] C.L. Blake, C.J. Merz, UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California at Irvine, California, 1998. Available from: http://www.ics.uci.edu/~mlearn/MLRepository.html, accessed 12 July 2008.

[2] F.M. Burnet, A modification of Jerne's theory of antibody production using the concept of clonal select, Australian Journal of Science 20 (1957), 67–69.

[3] D. Dasgupta, F. Nino, Immunological Computation: Theory and Applications, CRC Press, Florida, 2008.

[4] D. Dasgupta, S. Yu, F. Nino, Recent advances in artificial immune systems: models and applications. Applied Soft Computing 11 (2011), 1574–1581.

[5] L.N. De Castro, F.J. Von Zuben, Learning and optimization using the clonal selection principle, IEEE Transactions on Evolution Computation 6 (2002), 239–251.

[6] V.S. Desai, D.G. Conway, J.N. Crook, G.A. Overstreet Jr., Credit scoring models in the credit union environment using neural networks and genetic algorithms, IMA Journal of Mathematics Applied in Business and Industry 8 (1997), 323–346.

[7] V.S. Desai, J.N. Crook, G.A. Overstreet Jr., A comparison of neural networks and linear scoring models in the credit union environment, European Journal of Operational Research 95 (1996), 24–37.

[8] A.L. Dimitras, R. Slowinski, R. Susmaga, C. Zopounidis, Business failure prediction using rough sets, European Journal of Operational Research 114 (1999), 263–280.

[9] A.P. Engelbrecht, Computational Intelligence: An Introduction, 2nd ed., John Wiley &

Sons, West Sussex, 2007.

[10] S. Forrest, A.S. Perelson, L. Allen, R. Cherukuri, Self-nonself discrimination in a computer, in: Proceedings of the IEEE Symposium on Research in Security and Privacy, Oakland, 1994, pp. 202–212.

[11] J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd ed., Morgan Kaufmann Publishers, San Francisco, 2006.

[12] E. Hart, J. Timmis, Application areas of AIS: the past, present, and the future, Applied Soft Computing 8 (2008), 191–201.

[13] W.E. Henley, D.J. Hand, A k-nearest neighbor classifier for assessing consumer credit risk, Statistician 44 (1996), 77–95.

[14] C.-L. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, S. Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study, Decision Support Systems 37 (2004), 543–558.

[15] C.-L. Huang, M.-C. Chen, C.-J. Wang, Credit card scoring with a data mining approach based on support vector machine, Expert Systems with Applications 34 (2007), 847–856.

[16] N.K. Jerne, Towards a network theory of the immune system, Annals of Immunology 125C (1974), 373–389.

[17] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the International Joint Conference on Artificial Intelligence 1 (1995), 1137–1145.

[18] J. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, 1992.

[19] Y.D. Lan, J.G. Chen, G. Wets, Improving associative classification by incorporating novel interestingness measures, Expert Systems with Applications 31 (2006), 184–192.

[20] K. Leung, F. Cheong, C. Cheong, Generating compact classifier systems using simple

artificial immune system, IEEE Transactions on Systems, Man, and Cybernetics– Part B: Cybernetics 37 (2007), 1344–1356.

[21] R. Malhotra, D.K. Malhotra, Differentiating between good credits and bad credits using neuro-fuzzy systems, European Journal of Operational Research 136 (2002), 190–211.

[22] P. Matzinger, Tolerance, danger, and the extended family, Annual Review of Immunology 12 (1994), 991–1045.

[23] S. Piramuthu, Financial credit-risk evaluation with neural and neurofuzzy systems, European Journal of Operational Research 112 (1999), 310–321.

[24] A.S. Perelson, Immune network theory, Immunological Reviews 10 (1989), 5–36.

[25] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufman, California, 1993.

[26] A.K. Reichert, C.C. Cho, G.M. Wagner, An examination of the conceptual issues involved in developing credit-scoring models, Journal of Business and Economic Statistics 1 (1983), 101–114.

[27] J. Timmis, M. Neal, J. Hunt, An artificial immune system for data analysis, BioSystems 55 (2000), 143–150.

[28] J. Timmis, Artificial Immune Systems: A Novel Data Analysis Technique Inspired by the Immune Network Theory, Ph.D. Dissertation, Department of Computer Science, University of Wales, Aberystwyth, 2001.

[29] J. Timmis, M. Neal, A resource limited artificial immune system for data analysis, Knowledge-Based Systems, 14 (2001), 121–130.

[30] C.-F. Tsai, M.-L. Chen, Credit rating by hybrid machine learning techniques, Applied Soft Computing 10 (2010), 374–380.

[31] A. Watkins, AIRS: A Resource Limited Artificial Immune Classifier, M.S. Thesis, Mississippi State University, Starkville, 2001.

[32] A. Watkins, J. Timmis, L. Boggess, Artificial immune recognition system (AIRS): an

immune-inspired supervised learning algorithm. Genetic Programming and Evolvable Machines 5 (2004), 291–317.

[33] D. West, Neural network credit scoring models, Computers & Operations Research 27 (2000), 1131–152.

Highlights

> We propose an AINE-based classifier to evaluate the applicants' credit. >Our classifier outperforms the SAIS and SVM classifiers in credit scoring. >We compare our classifier with other AIS-based classifiers in the different datasets. > We show that our classifier can rival the AIRS-based classifiers.