



Consumer credit scoring using an artificial immune system algorithm

Leung Kan Hing, Kevin; Cheong, France; Cheong, Christopher

<https://researchrepository.rmit.edu.au/esploro/outputs/conferenceProceeding/Consumer-credit-scoring-using-an-artificial/9921860538401341/filesAndLinks?index=0>

Leung Kan Hing, K., Cheong, F., & Cheong, C. (2007). Consumer credit scoring using an artificial immune system algorithm. Proceedings of 2007 IEEE Congress on Evolutionary Computation, 3377–3384.
<https://doi.org/10.1109/CEC.2007.4424908>

Published Version: <https://doi.org/10.1109/CEC.2007.4424908>

Repository homepage: <https://researchrepository.rmit.edu.au>

© 2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Downloaded On 2024/04/28 10:47:27 +1000

Consumer Credit Scoring using an Artificial Immune System Algorithm

Kevin Leung, France Cheong, Christopher Cheong

Abstract—Credit scoring has become a very important task in the credit industry and its use has increased at a phenomenal speed through the mass issue of credit cards since the 1960s. This paper compares the performance of current classifiers against an artificial intelligence technique based on the natural immune system, named simple artificial immune system (SAIS). Experiments were performed on three benchmark credit datasets and SAIS was found to be a very competitive classifier.

I. INTRODUCTION

Credit scoring is one of the most successful applications and operations research techniques used in banking and finance, and is also one of the earliest financial risk management tools developed [1]. Its aim is to produce a score that any lending institution can use to classify applicants into two groups: one group which is credit-worthy and which is likely to repay its financial obligation and another group which is non-credit-worthy and whose application for credit will be rejected due to a high possibility of defaulting on its financial obligation. Credit scoring is therefore a typical classification problem.

Credit scoring was developed by Fair and Isaac in the early 1960s and the credit risk modeling literature has grown extensively since the seminal work by Altman [2] and Merton [3]. Indeed, since the 1960s, credit scoring has played a vital role in the phenomenal growth of consumer credit, especially for credit cards. It has been widely accepted in the United States of America in the early 1980s and United Kingdom in the early 1990s. The number of credit card owners has also increased rapidly in Australia. According to the Reserve Bank of Australia, the number of credit card accounts has increased from 8.1 million in 1998 to 10.4 million in 2003 [4]. As for the number of credit card transactions, it has increased by 160% from 394.3 million in 1998 to 1,026.0 million in 2003. However, as consumer credit increases at an extraordinary rate, so too have consumer bankruptcies. According to the Australian Government Inspector-General in Bankruptcy, the number of customers, including both business and non-business, who filed for bankruptcies has increased by more than 185% since 1988 [5].

Despite an increase in consumer bankruptcies, competition in the consumer loan market is getting more intense everyday. Lenders are now using different types of techniques to evaluate consumer loans in order to reduce loan losses [6]. More recently, artificial intelligence (AI) techniques like

K. Leung, F. Cheong and C. Cheong are with the School of Business IT, RMIT University, Melbourne, AUSTRALIA (email: {kevin.leung,france.cheong,chris.cheong}@rmit.edu.au).

expert systems and artificial neural networks (ANNs) have been used for building scorecards. Other techniques such as genetic algorithms (GAs) and k-Nearest Neighbour (kNN) have been tried without much success. They have not become popular because, although their forecasting abilities to classify applicants can perhaps equal to those of conventional statistical models, they do not seem to give any extra advantages [7]. ANNs, for instance, are commonly considered as black-box techniques without logic or rule explanation, i.e. the resulting solution is not easily interpretable. kNN, on the other hand, requires major systems investment since generating the nearest-neighbour rule is very computationally intensive.

A more recent form of AI technique, known as artificial immune system (AIS), is rapidly emerging. It is based on the natural immune system principles and it can offer strong and robust information processing capabilities for solving complex problems. Even though AIS has been used in the area of pattern recognition and classification, there has only been a single case where it has been applied to credit scoring purposes. Watkins et al. [8] found that their AIS, known as artificial immune recognition system (AIRS), exhibited the best performance of any single classifier used on their dataset.

It is important to continuously search for new techniques to improve the performance of scorecards as this is motivated by the fact that with the increasing volume of borrowing, even a small drop in bad debt can save millions of dollars. As such, this study will introduce the use of a new AIS classifier system in the context of credit scoring and compare its performance against current classifiers.

The rest of the paper is organized as follows. Section II discusses some related work on credit scoring AI techniques, whilst section III gives an explanation of the algorithm and implementation of the new classifier system. Section IV provides details of the tests performed and results obtained, and section V concludes the paper.

II. A REVIEW OF AI CREDIT SCORING TECHNIQUES

Biological systems are a rich source of metaphors for constructing intelligent information processing systems. These systems can be classified as: brain-nervous systems (artificial neural networks), genetic systems (genetic algorithms) and immune systems (artificial immune systems). Compared to ANNs and GAs, which have been widely applied to various fields, applications of AIS are relatively few.

A. Artificial Neural Networks (ANNs)

ANNs are inspired by the functionality of the nerve cells in the brain. Just like humans, ANNs can learn to recognise patterns by repeated exposure to many different examples. They are non-linear models that can classify based on pattern recognition capabilities [9]. This gives them an advantage over conventional statistical techniques used in industry which are primarily linear.

In the field of credit scoring, studies have shown that neural networks perform significantly better than statistical techniques such as discriminant analysis (DA) and logistic regression (LR) analysis [6, 10]. West [11] investigated the accuracy of quantitative models commonly used for credit scoring. He found that ANNs can improve the credit scoring accuracy and found LR to be the most accurate of the conventional methods used.

As mentioned in section I, ANNs solutions are not easily interpretable. Also they require extensive training and all these factors have limited their applications in the field of credit scoring.

B. Genetic Algorithms (GAs)

GAs are efficient problem-solving mechanisms that are inspired by the mechanisms of biological evolution [12–14]. The aim of GAs is to continuously evolve a problem's solution over many processing cycles, each time producing better solutions.

The use of GAs is now growing rapidly with successful applications in finance trading, fraud detection and other areas of credit risk. Desai et al. [15] investigated the use of GAs as a credit scoring model in a credit-union environment while Yobas et al. [16] compared the predictive performances of four techniques, one of which is GAs, in identifying good and bad credit card holders. Interestingly, they [16] found that DA performed better followed by GAs.

C. Artificial Immune Systems (AIS)

AIS is based on the natural immune system of the body. Just like ANNs, AIS can learn new information, recall previously learned information, and perform pattern recognition in a highly decentralised way [17].

The main study which regards AIS as a supervised classifier system was done by Watkins [18]. The classifier system was named AIRS and it is based on the principle of resource-limited AIS and made use of artificial recognition balls. AIRS has proved to be a very powerful classification tool and when compared to the 30 best classifiers on publicly available classification problem sets, one of which is a credit scoring dataset, it was found to be among the top five to eight classifiers for every problem set, except for one in which it ranked second [8].

D. Summary

The literature on credit scoring and the most common AI techniques used for building scorecards has been reviewed. Some studies [15, 16, 19] found statistical techniques to perform better than AI techniques, while others [20, 21]

concluded just the opposite. Their comparison results are shown in Table I. It should be noted that the numbers should be compared across the rows rather than between the rows since different datasets were used by each of the five different authors. Some of these results were obtained from Thomas et al.'s book [1].

TABLE I
COMPARISON OF CLASSIFICATION ACCURACY OF DIFFERENT CREDIT SCORING TECHNIQUES

| Authors | DA | LR | Decision Trees | Linear Prog. | ANNs | GAs |
|---------|-------|-------|----------------|--------------|-------|-------|
| [21] | 87.5% | 89.3% | 93.2% | 86.1% | - | - |
| [19] | 77.5% | - | 75.0% | 74.7% | - | - |
| [20] | 43.4% | 43.3% | 43.8% | - | - | - |
| [15] | 66.5% | 67.3% | - | - | 66.4% | - |
| [16] | 68.4% | - | 62.3% | - | 64.2% | 64.5% |

III. METHODOLOGY

This section presents an overview of the proposed algorithm and classifier system which is named simple artificial immune system (SAIS). As its name implies, SAIS is very simple in that it adopts only the concept of affinity maturation which deals with stimulation, cloning and mutation as opposed to currently available AIS which tend to focus on several particular subsets of the features found in the natural immune system. It also generates a compact classifier using only a predefined number of exemplars per class. This will be further discussed in the next section which also provides the pseudocode explaining how the SAIS model works.

A. Conventional AIS Algorithm

In a conventional AIS algorithm (such as [8]), a classifier system is constructed as a set of exemplars that can be used to classify a wide range of data and in the context of immunology, the exemplars are known as B-cells and the data to be classified as antigens. A typical AIS algorithm operates as follows:

- 1) First, a set of training data (antigens) is loaded and an initial classifier system is created as a pool of B-cells with attributes either initialised from random values or values taken from random samples of antigens.
- 2) Next, for each antigen in the training set, the B-cells in the cell pool are stimulated. The most highly stimulated B-cell is cloned and mutated, and the best mutant is inserted in the cell pool. To prevent the cell pool from growing to huge proportions, B-cells that are similar to each other and those with the least stimulation levels are removed from the cell pool.
- 3) The final B-cell pool represents the classifier.

The conventional AIS algorithm is shown in Algorithm 1. From the description of the algorithm, three problems are apparent with conventional AIS algorithms:

- 1) Only one pass through the training data does not guarantee the generation of an optimal classifier.

- 2) Finding optimal B-cells does not guarantee the generation of an optimal classifier as local optimizations at the B-cell level does not necessarily imply global optimization at the B-cell pool level.
- 3) The simple population control mechanism of removing duplicates cannot guarantee a compact B-cell pool size. Many of the early AIS classifiers reported in the literature [22, 23] suffer from the problem of huge size. Good B-cells may be lost during the removal process. A conventional AIS classifier was experimented with and the size of the cell pool was found to grow to astronomical proportions when using such a simple population control mechanism.

Algorithm 1 Conventional AIS Algorithm

```

Load antigen population {training data}
Generate pool of B-cells with random values or values
from random antigens
for each antigen in population do
  Present antigen to B-cell pool
  Calculate stimulation level of B-cells
  Select most highly stimulated B-cell
  if stimulation level > threshold then
    Clone and mutate selected B-cell
    Select best mutants and insert into B-cell pool
  end if
  Delete similar and least stimulated B-cells from B-cell
  pool
end for
Classifier ← B-cell pool
  
```

B. SAIS Algorithm

In order to address the issues present in conventional AIS algorithms, the SAIS algorithm is designed to operate as follows:

- 1) First, a set of training data (antigens) is loaded and an initial classifier system is created as a single B-cell containing a predefined number of exemplars initialized from random values. The purpose and content of this B-cell is different from the one used in conventional AIS algorithms. This B-cell represents the complete classifier and it contains one or more exemplars per class to classify. A B-cell in a conventional AIS algorithm, however, represents exactly one exemplar and the complete classifier is made up of a pool of B-cells.
- 2) Next, an evolution process is performed and iterated until the best possible classifier is obtained. The current B-cell is cloned and the number of clones that can be produced is determined by the clonal rate and hypermutation rate. Mutants are then generated by using the hypermutation process found in natural immune systems. More specifically, this is achieved by randomly mutating the attributes of each clone created and storing them in a 3-dimensional array. Such an

array is used because it is easier to store the attributes, classes and exemplars [24].

- 3) Each mutant is then evaluated by using the classification performance. The classification performance is a measure of the percentage of correctly classified data. If the classification performance of the best mutant is better than that of the current B-cell, then the best mutant is taken as the current B-cell. The measure of stimulation is different from one used in conventional systems in that a classification performance is used as a measure of stimulation of the complete classifier on all the training data rather than the distance (or affinity) between part of the classifier (a B-cell) and part of the data (an antigen).
- 4) The current B-cell represents the classifier.

The SAIS algorithm is shown in Algorithm 2. Using a B-cell to represent the whole classifier rather than part of the classifier has several advantages:

- 1) Optimizations are performed globally rather than locally and nothing gets lost in the evolution process.
- 2) There is no need for any population control mechanism as the classifier consists of a small predefined number of exemplars. So far in the experiments performed, only one exemplar per class to be classified was used. This ensures the generation of the most compact classifier possible.

Algorithm 2 SAIS Algorithm

```

Load antigen population {training data}
Current B-cell ← randomly initialized B-cell
repeat
  Evolve the B-cell by cloning and mutation
  Evaluate mutated B-cells by calculating their classifica-
  tion performance
  New B-cell ← mutated B-cell with best performance
  if performance of new B-cell > current B-cell then
    Current B-cell ← new B-cell
  end if
until maxIteration
Classifier ← current B-cell
  
```

A diagram showing the differences between a conventional AIS and our SAIS is provided in Figure 1.

C. Model Implementation

SAIS was implemented in Java using the Repast¹ agent-based modelling framework. A minimum distance classification method, which has a linear computational complexity, was used. It is an exemplar-based method where the numbers of attributes of a single exemplar per class are stored in the classifier. If there are two classes for instance, the complete classifier will consist of two exemplars and their attributes. This method is explained in more detail in the following section.

¹Available from <http://repast.sourceforge.net>

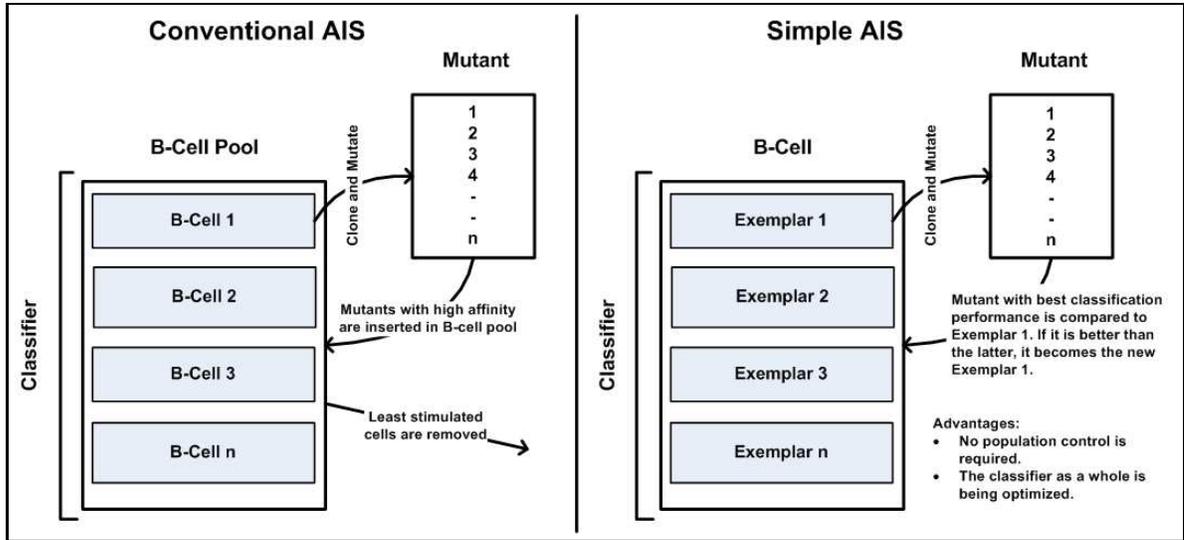


Fig. 1. Comparison of Conventional AIS and Simple AIS

1) *Minimum Distance Classification Method* : In this exemplar-based method, a distance measure is used to classify the data. This approach is adapted from instance-based learning (IBL) [25] which is a learning paradigm in which algorithms store the training data and use a distance function to classify the data to be tested. The heterogeneous Euclidean-overlap metric (HEOM) [26] is used. It can handle both categorical and continuous attributes and is defined as:

$$totalDist(x_1, x_2) = \sqrt{\sum_{i=1}^n dist(x_{1,i}, x_{2,i})^2} \quad (1)$$

where x_1 is an exemplar, x_2 is an antigen and n is the number of attributes. The distance between an exemplar and one antigen is calculated as:

$$dist(x_{1,i}, x_{2,i}) = \begin{cases} 1, & \text{if missing} \\ catDist(x_{1,i}, x_{2,i}), & \text{if categorical} \\ contDist(x_{1,i}, x_{2,i}), & \text{if continuous} \end{cases} \quad (2)$$

Missing attributes are handled by returning a distance of one. This is because the smaller the distance between an antigen and an exemplar, the more likely the antigen will be classified in the class of that particular exemplar. Also, since all the data has been normalised and their values range between zero and one, a distance of one, which is the maximum possible instance to any attribute, is allocated to each missing attribute. The data has been normalized in order to avoid the problem of overpowering the other attributes if one of them has a relatively large range.

Categorical attributes are handled by the overlap function:

$$catDist(x_{1,i}, x_{2,i}) = \begin{cases} 0, & \text{if } x_{1,i} = x_{2,i} \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

while continuous attributes are handled by the Euclidean function which is calculated as:

$$contDist(x_{1,i}, x_{2,i}) = (x_{1,i} - x_{2,i}) \quad (4)$$

The minimum distance is then chosen to determine the class to classify each antigen. The predicted classifications are then checked against the testing data and the percentage of correctly classified data can thus be generated.

D. System Parameters

The system parameters used by the SAIS classifier are shown in Table II.

TABLE II
SYSTEM PARAMETERS OF SAIS

| Name | Description and value |
|--------------------------|--|
| <i>clonalRate</i> | Default value = 10. |
| <i>hyperMutationRate</i> | Default value = 10. Number of clones that can be mutated = 100 (<i>clonalRate</i> × <i>hyperMutationRate</i>). |
| <i>maxIterations</i> | 600 iterations were enough for the performance of the classifier to become constant. In fact, an average of 224 iterations were enough for the performance of SAIS to become constant. |
| <i>probMutation</i> | Probability of mutation = 0.7. |

clonalRate: An integer value that is used to determine the number of mutated clones an exemplar is allowed to produce

hyperMutationRate: An integer value that is used to determine the number of mutated clones that are generated into the cell population

maxIterations: Maximum number of iterations

probMutation: Probability that a given clone will mutate

Readers are referred to [24, 27] for additional information on SAIS.

IV. DATA ANALYSIS

The classification performance of SAIS was tested on three consumer credit datasets. One of them was obtained from Thomas et al. [1], while the other two were obtained from the University of California Irvine [28]. The last two datasets are publicly available benchmark datasets, known as the Australian and German Credit Approval datasets, and they were also used in the *Statlog* project [29]. The results of the Australian and German datasets will be compared against those obtained from the *Statlog* project and also against AIRS's results, which were generated using its default settings. Table III shows the description of the three datasets. It should be noted that 'con' stands for continuous and 'cat' stands for categorical attributes.

TABLE III
DATASETS USED FOR EXPERIMENTS

| Dataset | Attribute type | n | Classes | Missing attributes |
|------------|----------------|------|----------|--------------------|
| Australian | 6 con | 690 | 307 good | 37 |
| | 9 cat | | 383 bad | |
| German | 7 con | 1000 | 700 good | - |
| | 13 cat | | 300 bad | |
| Thomas | 10 con | 1225 | 902 good | - |
| | 4 cat | | 323 bad | |

A. Experiment

A stepwise regression analysis was performed on the three datasets in order to select the most relevant explanatory attributes. This regression method is essentially a forward selection procedure, coupled with the possibility of removing a variable, just as in a backward elimination procedure [30]. A full list of the independent variables of the three datasets used in this study after data pre-processing is shown in Table IV. While it was possible to get descriptive information on the attributes used in the German and Thomas datasets, it was not possible to do so for the Australian dataset due to confidentiality issues. The adjusted R^2 for each dataset has also been included. The Australian dataset has the highest adjusted R^2 , meaning that its predictors are more able to explain the dependent variable.

To be comparable with other classifiers used in the literature [31], a 10-fold cross validation (CV) technique was used to partition each dataset into training and testing sets. 10 different sets of data, each containing one portion as the testing set and nine portions as the training set, were therefore generated. SAIS was run 600 times on the 10 training sets of each dataset and results show that the performance of the classifier becomes constant after an average of 224 iterations.

The classifier was then run on the 10 testing sets of each dataset, with each set of data producing a classification performance of SAIS. The 10 classification results were averaged to yield an overall classification performance of the model. Due to the fact that SAIS is evolutionary and the results obtained are unlikely to be similar twice, i.e.

TABLE IV
ATTRIBUTES USED FOR EXPERIMENTS

| Australian dataset | German dataset | Thomas dataset |
|------------------------|---|------------------------------|
| A2 | Status of checking account | Year of birth |
| A3 | Duration | Number of dependents |
| A4 | Credit history | Home phone |
| A5 | Credit amount | Spouse's income |
| A6 | Saving account bonds | Applicant's income |
| A8 | Present employment since | Residential status |
| A9 | Installment rate in percentage of disposable income | Mortgage balance outstanding |
| A10 | Personal status and sex | Outgoings on loans |
| A11 | Other debtors/guarantors | Outgoings on hire purchase |
| A12 | Property | Outgoings on credit cards |
| A14 | Other installment plans | |
| A15 | Housing | |
| | Number of existing credits at this bank | |
| | Telephone | |
| | Foreign worker | |
| adjusted $R^2 = 0.594$ | adjusted $R^2 = 0.227$ | adjusted $R^2 = 0.058$ |

SAIS is non-deterministic, the experiment described above was performed 10 times, i.e. 10×10 -fold CV. The results obtained were again averaged.

B. Performance Measure

The receiver operating characteristics (ROC) curve and Gini coefficient (G), which can be calculated from the area under the ROC curve (AUC) (refer to Equation (5)) [32], would have been the most appropriate measures of model performance in this study. ROC curves have become the standard tool for assessing the accuracy of model predictions in the field of medical diagnosis and are now becoming increasingly used in the machine learning and financial environment [33]. G is also the measure of performance that is being used by financial institutions in the field of credit scoring.

$$AUC = \frac{1}{2}(G + 1) \quad (5)$$

Since SAIS is a discrete classifier and can only produce a class decision (i.e. a good or a bad) as result on each instance, ROC curves and hence G cannot be generated. This is because when such a discrete classifier is applied to the testing data, it produces a single confusion matrix (see Figure 2), which in turn corresponds to a single ROC point [33].

It should, however, be noted that while ROC curves cannot be generated, ROC graphs can be obtained. These are 2-dimensional graphs in which the TP rate (refer to Equation (6)) and FP rate (refer to Equation (7)) are plotted on the y- and x- axes respectively. They give the trade-offs between benefits (TP) and costs (FP). A ROC graph will be plotted for

| | | | |
|------------------------|----------|--------------------|--------------------|
| | | True class | |
| | | g | b |
| Predicted class | g | True Positives | False Positives |
| | b | False Negatives | True Negatives |
| Total: | | G | B |

Fig. 2. Confusion Matrix

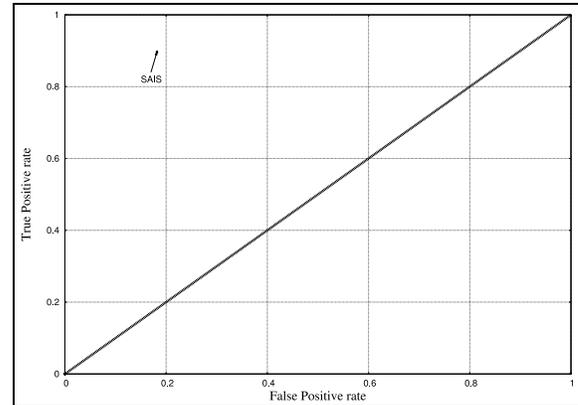


Fig. 3. ROC graph for Australian dataset

each dataset. The single ROC point of each classifier will be obtained by averaging all the TP and FP rates of each testing dataset.

$$tprate = \frac{TP}{G} \quad (6)$$

$$fprate = \frac{FP}{B} \quad (7)$$

Another performance measure used is the classification accuracy. The latter is the percentage of correctly classified good and bad classes. The classification accuracy is used due to the fact that other researchers, who have worked with these three datasets, have used it as their main measure of performance. In order to compare the performance of SAIS against other classifiers, this performance measure has to be used.

$$accuracy = \frac{TP + TN}{G + B} \times 100\% \quad (8)$$

C. Results

1) *Australian Dataset* : Based on the experiments performed, it has been found that for the Australian dataset, which has a near equal distribution of good and bad classes, SAIS exhibits a classification performance of 85.2% with a standard deviation (SD) of 0.2. The ROC graph (see Figure 3) also indicates that SAIS is a good classifier since it is close to the point (0,1), which represents perfect classification.

As mentioned previously, the results of this study are compared against those obtained from the *Statlog* project. The results were obtained from [29,31]. Table V shows the percentage accuracy of the different classifiers when used on the Australian dataset. The results show that SAIS is ranked sixth with a difference in percentage of only 1.7% from the best model, indicating that it is a very competitive classifier.

2) *German Dataset* : A 75.4% classification accuracy with a SD of 0.6 was obtained when SAIS was applied on the German dataset. The ROC graph (see Figure 4) shows that SAIS is 'conservative' meaning that it makes positive classifications only with strong evidence. It therefore makes few FP errors, but it also very often has few TP rates [33].

TABLE V
COMPARATIVE RESULTS FOR AUSTRALIAN DATASET

| Rank | Classifier | Accuracy | From |
|------|-------------|-------------|------------|
| 1 | Cal5 | 86.9% | |
| 2 | ITrule | 86.3% | |
| 3 | DIPOL92 | 85.9% | |
| 4 | CART | 85.5% | |
| 5 | RBF | 85.5% | |
| 6 | SAIS | 85.2% (0.2) | This study |
| 7 | AIRS | 85.2% (5.6) | This study |
| 8 | CASTLE | 85.2% | |
| 9 | Naive Bayes | 84.9% | |
| 10 | IndCART | 84.8% | |
| 11 | Backprop | 84.6% | |
| 12 | C4.5 | 84.5% | |
| 13 | SMART | 84.2% | |
| 14 | Baytree | 82.9% | |
| 15 | k-NN | 81.9% | |
| 16 | NewID | 81.9% | |
| 17 | Acsquare | 81.9% | |
| 18 | LVQ | 80.3% | |
| 19 | ALLOC80 | 79.9% | |
| 20 | CN2 | 79.6% | |
| 21 | Quadisc | 79.3% | |
| 22 | Default | 56.0% | |

Similar to what was done for the Australian dataset, the results obtained are compared against those from the *Statlog* project. However in this particular dataset, the *Statlog* project results are associated with cost whereby the cost in classifying a bad debtor as good is five times more costly than the opposite. The results of this study were therefore converted to the average cost. This was achieved by multiplying the confusion matrix by the cost matrix, summing the entries and dividing by the number of observations [29].

The results obtained are shown in Table VI. SAIS is ranked third with an average cost of 59%. Such a low cost was obtained for SAIS because the latter is considered as 'conservative' and as such, has low FP rate. The results in Table VI again proves that SAIS is a very competitive classifier.

Some recent accuracy performance measure results for this dataset were also obtained from the literature [34, 35]. These

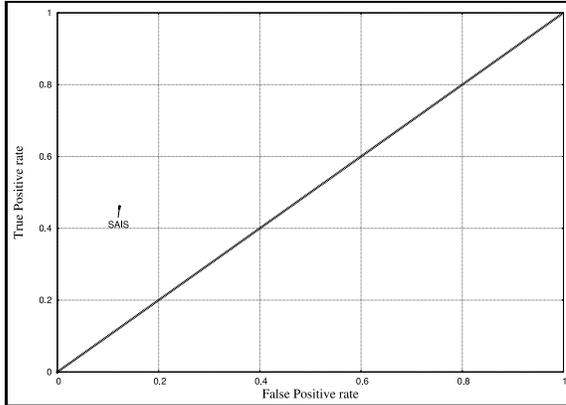


Fig. 4. ROC graph for German dataset

TABLE VI
COMPARATIVE RESULTS FOR GERMAN DATASET

| Rank | Classifier | Cost | From |
|------|-------------|-------|------------|
| 1 | CASTLE | 0.583 | |
| 2 | ALLOC80 | 0.584 | |
| 3 | SAIS | 0.590 | This study |
| 4 | DIPOL92 | 0.599 | |
| 5 | SMART | 0.601 | |
| 6 | Cal5 | 0.603 | |
| 7 | CART | 0.613 | |
| 8 | Quadisc | 0.619 | |
| 9 | k-NN | 0.694 | |
| 10 | Default | 0.700 | |
| 11 | Naive Bayes | 0.703 | |
| 12 | IndCART | 0.761 | |
| 13 | Backprop | 0.772 | |
| 14 | Baytree | 0.778 | |
| 15 | CN2 | 0.856 | |
| 16 | AC | 0.878 | |
| 17 | ITrule | 0.879 | |
| 18 | NewID | 0.925 | |
| 19 | LVQ | 0.963 | |
| 20 | RBF | 0.971 | |
| 21 | C4.5 | 0.985 | |
| 22 | Kohonen | 1.160 | |

accuracies are shown in Table VII.

The results show that the NN algorithm by Kim and Sohn [34] recorded the highest accuracy, which is 2.6% higher than that of SAIS. However, since its confusion matrix was provided and if its average cost was calculated in the same way and then included in Table VI for comparison purposes, it would have been ranked twelfth, at a cost of 74%. This leads to the conclusion that accuracy is a very misleading performance measure when it comes to unbalanced dataset.

It is interesting to note that while the accuracy is a misleading performance measure for unbalanced datasets, nevertheless most studies still use it. This study also made use of accuracy as one of its performance measures; however, this was primarily done for comparison purposes.

3) *Thomas Dataset* : SAIS recorded a classification accuracy of 74.3% with a SD of 0.3 and the findings are very similar to those obtained for the German dataset in

TABLE VII
ACCURACY RESULTS FOR GERMAN DATASET

| Rank | Model | Accuracy | From |
|------|-------------|-------------|------------|
| 1 | NN | 78.0% | [34] |
| 2 | SAIS | 75.4% (0.6) | This study |
| 3 | Naive Bayes | 74.7% | [35] |
| 4 | CBA | 74.4% | [35] |
| 5 | C4.5 | 72.4% | [35] |
| 6 | AIRS | 71.3% (4.6) | This study |

that SAIS is a ‘conservative’ classifier (see Figure 5). The ROC graph also indicates that the model is far from being a good classifier since its ROC point are far from point (0,1). However, it would be wrong to suggest that it is a bad classifier since it performed well for the Australian and German datasets. The main reason for such a behaviour can be explained by the adjusted R^2 (see Table IV). Since the Thomas dataset has a very low adjusted R^2 compared to the Australian dataset, most of its attributes are less likely to explain the dependent variable. This indicates that from the data available in the Thomas dataset, it was hard for SAIS to predict accurately, thereby explaining why such a ROC graph was obtained.

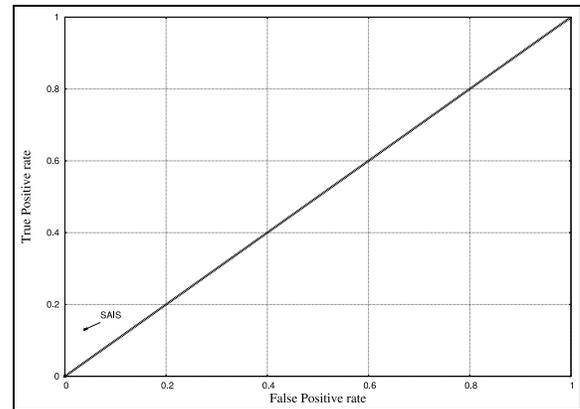


Fig. 5. ROC graph for Thomas dataset

Table VIII shows the accuracy of other classifiers which have used the Thomas dataset. Readers should be aware that few studies made use of this dataset, probably because it is not publicly available. Again the results show that SAIS is a competitive classifier, being ranked second.

4) *Summary* : Based on the above analysis, it can be said that SAIS is a very competitive classifier, being among the top five classifiers for the German and Thomas datasets and being ranked sixth for the Australian dataset.

V. CONCLUSION AND FUTURE WORK

Most real credit scoring datasets are unbalanced datasets. However, what goes in the training dataset remains the decision of the credit analyst and the financial institution. In this study, both balanced and unbalanced datasets were

TABLE VIII
COMPARATIVE RESULTS FOR THOMAS DATASET

| Rank | Classifier | Accuracy | From |
|------|------------|-------------|------------|
| 1 | LSSVM | 89.2% | [36] |
| 2 | SAIS | 74.3% (0.3) | This study |
| 3 | B-FSVM | 66.2% | [37] |
| 4 | AIRS | 65.9% (4.5) | This study |
| 5 | U-FSVM | 65.4% | [37] |
| 6 | SVM | 65.4% | [37] |
| 7 | LR | 64.1% | [37] |
| 8 | NN | 62.1% | [37] |

used. A new and simple AIS algorithm and classifier was implemented and the performance of SAIS was tested on three different datasets. It was found that SAIS is a very competitive classifier.

Future work lies in improving the performance of SAIS by using multiple exemplars per class, instead of one exemplar which was used in this study. There is also the intention of generating a score for each instance so that a ROC curve and hence G can be obtained. Using a GA to automatically select the most relevant attributes of a dataset can also be used. Finally, testing the model on a real consumer credit dataset, which can be obtained from a leading financial institution, is also envisaged.

REFERENCES

- [1] L. C. Thomas, D. B. Edelman, and J. N. Crook, *Credit Scoring and Its Applications*, p. k. sen ed. North-Holland, Amsterdam: Elsevier Science Publishers, 2002.
- [2] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *Journal of Finance*, pp. 589–609, September 1968.
- [3] R. C. Merton, "On the pricing of corporate debt: the risk structure of interest rates," *Journal of Finance*, vol. 29, pp. 449–470, 1974.
- [4] Euromonitor, "Financial Cards in Australia," 2004, http://www.euromonitor.com/Financial_Cards_in_Australia.
- [5] Australian Government, "Australian Government Inspector-General in Bankruptcy," 2004, <http://www.ag.gov.au>.
- [6] R. Malhotra and D. K. Malhotra, "Evaluating consumer loans using neural networks," *The International Journal of Management Science*, vol. 31, pp. 83–96, 2003.
- [7] A. Lucas, "Statistical challenges in credit card issuing," *Applied Stochastic Models in Business and Industry*, vol. 17, pp. 83–92, 2001.
- [8] A. Watkins, J. Timmis, and L. Boggess, "Artificial immune recognition system (AIRS): An immune-inspired supervised learning algorithm," *Genetic Programming and Evolvable Machines*, vol. 5, pp. 291–317, 2004.
- [9] S. Goonatilake and P. Treleaven, *Intelligent Systems for Finance and Business*. New York: Wiley, 1995.
- [10] V. S. Desai, J. N. Crook, and G. A. Overstreet Jr., "A comparison of neural networks and linear scoring models in the credit union environment," *European Journal of Operational Research*, vol. 95, pp. 24–37, 1996.
- [11] D. West, "Neural network credit scoring models," *Computers & Operations Research*, vol. 27, pp. 1131–1152, 2000.
- [12] L. Davis, *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.
- [13] G. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. MA: Addison-Wesley, 1989.
- [14] J. H. Holland, "Genetic algorithms and classifier systems: foundations and future directions," in *Proc. 2nd International Conference on Genetic Algorithms*. Lawrence Erlbaum Associates, Inc., 1987, pp. 82–89.
- [15] V. S. Desai, D. G. Conway, J. N. Crook, and 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*, vol. 8, pp. 323–346, 1997.
- [16] M. B. Yobas, J. N. Crook, and P. Ross, "Credit scoring using neural and evolutionary techniques," *IMA Journal of Mathematics Applied in Business and Industry*, vol. 11, pp. 111–125, 2000.
- [17] A. Tarakanov and D. Dasgupta, "A formal model of an artificial immune system," *BioSystems*, vol. 55, pp. 155–158, 2000.
- [18] A. Watkins, "AIRS: A Resource Limited Artificial Immune Classifier," Master's thesis, Mississippi State University, December 2001, available at <http://www.cse.msstate.edu/~andrew/research/publications.html>.
- [19] M. Boyle, J. N. Crook, R. Hamilton, and L. C. Thomas, "Methods for credit scoring applied to slow payers," in *Credit Scoring and Credit Control*, L. C. Thomas, J. N. Crook, and D. B. Edelman, Eds. Oxford: Oxford University Press, 1992, pp. 75–90.
- [20] W. E. Henley, "Statistical Aspects of Credit Scoring," Ph.D. dissertation, The Open University, Milton Keynes, UK, 1995.
- [21] V. Srinivasan and Y. H. Kim, "Credit granting: A comparative analysis of classification procedures," *Journal of Finance*, vol. 42, pp. 665–681, 1987.
- [22] J. Timmis and M. Neal, "A resource limited artificial immune system for data analysis," *Knowledge-Based Systems*, vol. 14, pp. 121–130, 2001.
- [23] O. Nasraoui, D. Dasgupta, and F. Gonzalez, "A novel artificial immune system approach to robust data mining," in *Proc. Genetic and Evolutionary Computation Conference (GECCO)*, New York, July 2002, pp. 356–363.
- [24] K. Leung, F. Cheong, and C. Cheong, "Generating compact classifier systems using a simple artificial immune system," *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 2007, submitted for publication.
- [25] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, pp. 37–66, 1991.
- [26] D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, pp. 1–34, 1997.
- [27] K. Leung and F. Cheong, "A simple artificial immune system (SAIS) for generating classifier systems," in *Proc. AI 2006: Advances in Artificial Intelligence*, ser. Lecture Notes in Artificial Intelligence, vol. 4304. Berlin: Springer, 2006, pp. 151–160.
- [28] C. L. Blake and C. J. Merz, "UCI Repository of Machine Learning Databases," 1998, <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [29] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, *Machine Learning, Neural and Statistical Classification*. New York: Ellis Horwood, 1994.
- [30] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting Methods and Applications*, 3rd ed. John Wiley & Sons, 1998.
- [31] W. Duch, "Datasets used for Classification: Comparison of Results," 2000, <http://www.phys.uni.torun.pl/kmk/projects/datasets.html>.
- [32] D. J. Hand and R. J. Till, "A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems," *Machine Learning*, vol. 45, no. 2, pp. 171–186, 2001.
- [33] T. Fawcett, "ROC graphs: Notes and practical considerations for data mining researchers," Intelligent Enterprise Technologies Laboratory, HP Laboratories Palo Alto, Tech. Rep. HPL-2003-4, 2003.
- [34] Y. S. Kim and S. Y. Sohn, "Managing loan customers using misclassification patterns of credit scoring model," *Expert Systems with Applications*, vol. 26, pp. 567–573, 2004.
- [35] Y. Lan, D. Janssens, G. Chen, and G. Wets, "Improving associative classification by incorporating novel interestingness measures," *Expert Systems with Applications*, vol. 31, pp. 184–192, 2006.
- [36] K. K. Lai, L. Yu, L. Zhou, and S. Wang, "Credit risk evaluation with least square support vector machine," in *Proc. RSKT 2006*, ser. Lecture Notes in Artificial Intelligence, vol. 4062. Berlin: Springer, 2006, pp. 490–495.
- [37] Y. Wang, S. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 6, pp. 820–831, 2005.