Applied Soft Computing 12 (2012) 611–618

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

An artificial immune classifier for credit scoring analysis

Shiow-Yun Chang*, Tsung-Yuan Yeh

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

ARTICLE INFO

Article history:
Received 21 September 2009
Received in revised form 9 August 2011
Accepted 1 November 2011
Available online 12 November 2011

Keywords:
Artificial immune network
Classifier
Credit scoring
Data mining
Classification

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.

© 2011 Elsevier B.V. All rights reserved.

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

* Corresponding author. Tel.: +886 6 2757575; fax: +886 6 2362162.
E-mail address: shiowyun@mail.ncku.edu.tw (S.-Y. Chang).

1568-4946/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2011.11.002

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 and 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, Naïve-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. Section 5 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

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات