Neurocomputing 183 (2016) 70–78

Contents lists available at ScienceDirect

Neurocomputing

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

Financial credit analysis via a clustering weightless neural classifier

Douglas O. Cardoso [a], Danilo S. Carvalho [a], Daniel S.F. Alves [a], Diego F.P. Souza [a,*], Hugo C.C. Carneiro [a], Carlos E. Pedreira [a], Priscila M.V. Lima [b], Felipe M.G. França [a]

[a] PESC/COPPE - Universidade Federal do Rio de Janeiro - UFRJ, Brazil
[b] NCE, Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 31 August 2014
Received in revised form
5 April 2015
Accepted 2 June 2015
Available online 9 January 2016

Keywords:
Bleaching
ClusWiSARD
Clustering
Concept drifting
Credit assignment
Online learning

ABSTRACT

Credit analysis is a real-world classification problem where it is quite common to find datasets with a large amount of noisy data. State-of-the-art classifiers that employ error minimisation techniques, on the other hand, require a long time to converge, in order to achieve robustness. This paper explores ClusWiSARD, a clustering customisation of the WiSARD weightless neural network model, applied to two different credit analysis real-world problems. Experimental evidence shows that ClusWiSARD is very competitive with Support Vector Machine (SVM) w.r.t. accuracy, with the advantage of being capable of online learning. ClusWiSARD outperforms SVM in training time, by two orders of magnitude, and is slightly faster in test time.

1. Introduction

Credit analysis represents the complex tasks of deciding which credit applicants present a good probability of returning the granted credit and which do not. This task depends on many different factors, such as economic and cultural circumstances, and is often delegated to human experts. Human judgement, however, may not use explicit rules that can be referenced as basis for decision making. That could lead to conflicting analysis of the same problem instance from different experts. In some countries, this is considered illegal. This question would justify the design of a machine learning system that is able to replace the decisions of experts, providing a single analysis standard.

Important pattern recognition challenges can be found in credit analysis. For example, data can be noisy or corrupted due to problems in data collection. Data could also embed temporal information, possibly useful to identify *concept drift*: movement of populations, changes in economy, natural catastrophes [1], general news [2], etc. These and other factors may affect the relations pertinent to credit assignment. Class imbalance is also expected, as credit applications labeled as "good" are more frequent than "bad" ones.

How observations were gathered and labeled is also noteworthy. Labelling could be done *a priori*, according to a risk appraisal system already in use. Alternatively, this could be performed after observing if payment of granted requests was duly realised. A system trained with data from the first case aims to reproduce the behaviour of the established classification system, instead of attempting to excel it. In the second case, training data is the product of a filtering process, implying in a reduction of information about the population.

Different machine learning techniques have been analysed in the context of this problem. As discriminated by Tsai [3], they may be classified in three smaller sets, which are: single classifiers, classifier ensemble and hybrid classifiers. The first one contains single supervised models, like Support Vector Machine (SVM) [4–8], Multilayer-Perceptron (MLP) [9,4,10], Decision Trees (DT) [11] and Genetic Algorithm/Programming (GA/GP) [12,13]. Regarding classification accuracy over the UCI dataset, which was also used in this work, some results obtained were 77.34% by Ong, Huang and Tzeng [13] with the use of GP and 77.09%/76.59% by Tsai [4] with SVM and MLP, respectively. These models have achieved at most an accuracy of 77.34% working as single classifiers. However, they may achieve much better results when grouped together, forming classifier ensembles. For instance, Ghodselahi [14] has obtained 81.42% with the use of a SVMs ensemble, and Hoffmann [15] has reached 84.90% with a GA-based SVM. Some other approaches used GA-based MLP [16] and GA-based SVM [17] in other financial credit analysis datasets. The third category, called Hybrid Classifiers, contains approaches mixing two or more techniques. For instance, combining clustering and single classifiers. Previously a work [3] compared many different approaches

* Corresponding author.
E-mail address: diegofpsouza@gmail.com (D.F.P. Souza).

http://dx.doi.org/10.1016/j.neucom.2015.06.105
0925-2312/

and claimed to achieve up to 88.93% through the use of clustering and SVM. Since ClusWiSARD is an intrinsic clustering WiSARD classifier, it retains the single classifier characteristic. Therefore, its results are comparable with other single classifier approaches. Besides, this model could later be further improved through the use of other techniques, like ensemble learning and GA.

Financial institutions may lend money to different types of entities (people, businesses, non-profits, among others) and under various sets of conditions. This makes the context of credit analysis very diverse, with specialized methods in constant development for each category of credit operation. This work focuses on the case of retail credit, i.e., when the credit is given directly to the consumer (a person) rather than an organization. Retail credit analysis is mostly done by means of *credit scoring* [18], which has been the standard practice for decision making in credit risk management for the past 35 years in the US, UK and some other countries [18].

The purpose of credit scoring revolves around the ability to rank a prospective customer in a given set of categories, in order to assess the probability of timely payment of the debt over a period of time. As a result, a "grade" is given to the customer, which will serve to classify him or her as being a probable good or bad payer and thus facilitating the decision of giving the credit or not. Such "grades" are dependent on the method used for scoring, usually being an number but "good" and "bad" categories are also often used. The scoring methods are usually divided into two categories: statistical and non-statistical. The first category includes techniques such as logistic regression and discriminant analysis, with uses dating back from the beginning of credit scoring activities. The second category includes a wide range of computer backed algorithms such as neural networks, linear programming and genetic algorithms. Advances in machine learning research and computational capabilities over the past decades have promoted significant increases in predictive accuracy for many non-statistical methods, boosting their adoption, although use of statistical methods has not been abandoned [19]. More sophisticated non-statistical methods, such as ensembles, have shown a boost in adoption after the 2008 financial crisis, which resulted in restrictions regarding retail credit provision [20]. The method presented in this paper is non-statistical and is compared to another method of the same category for illustration of its capabilities. The method chosen for comparison is the support vector machine (SVM) which is often used for the scoring task.

Having an automated learning and classification mechanism that could offer a more precise solution is an attractive idea. It must be able to analyse vast amounts of data on credit applications and consider subtle relations between the actual financial data and the borrower profile. However, such mechanism would also need to be both efficient and robust in order to account for changes in the circumstances and sample biasing. Two classifying mechanisms which have potential to exhibit these characteristics are the WiSARD [21] weightless artificial neural network model and the Support Vector Machine (SVM) [22], which are introduced, respectively, in Sections 2.1 and 2.2. This work proposes the application of WiSARD weightless neural network model for the credit analysis problem, both in its traditional form, targeting simpler scenarios, as well as in a clustering oriented architecture, called ClusWiSARD, in order to deal with more complex ones. Preprocessing methods for the data analysis are also discussed. For comparison purposes, the same data is classified by a Support Vector Machine.

This paper is organized as follows. Section 2 presents the methods and materials used for this research, including classifier models (Sections 2.1 and 2.2) as well as data set handling (Section 2.3); Section 3 details metrics and implementation of the experiments, shows their results and discusses some of the interesting findings; Section 4 concludes this work by summarising the findings and present possible avenues for further work.

2. Methodology

2.1. WiSARD

A Weightless Artificial Neural Network (WANN) is a pattern recognition system whose main difference from other learning methodologies lies on the direct use of information storage in Random Access Memories (RAMs) [21]. No error minimisation technique is used. WANN operation uses the input to build a set of addresses to access RAM nodes contents.

This work adopts WiSARD (Wilkie Stonham and Aleksander Recognition Device) [21], a pioneering WANN architecture that is composed by distinct sets of RAM nodes called *discriminators*. Each discriminator is assigned to one of the classes of patterns to be recognized, i.e., the number of discriminators in the WiSARD network is the same as the number of classes. A discriminator consists of a single layer of RAM nodes, which are all initialised with the default value zero (0) in every addressable position. The network has also been extended with a tie breaking capability, called *bleaching* [23], in order to deal with inconclusive pattern classifications. With respect to this work, WiSARD speed was very useful: all its operations have polylogarithmic complexity on the number of input observations. Additionally, that only requires a small number of parameters to be set.

2.1.1. Input encoding

WiSARD is, originally, a Boolean neural network, so any input given to the architecture must be converted into a binary string. However, the most common description of data is by numerical and/or categorical attributes. A data conversion process must be applied to bridge this gap. This process may not be straightforward, as the similarity between any two observations should be preserved in the new representation. The preferred binary encodings for numeric features are the ones with a *Hamming distance* related to the numeric distance. Encodings which do not have this characteristic, e.g., IEEE 754 [24], should be avoided.

This conversion can be tuned with respect to a number of factors, such as domain knowledge and classifier performance on tests. After the conversion is made, the input is shuffled according to a fixed pseudorandom mask (defined at the creation of the network) and split to generate input addresses of all RAM nodes. During the training phase, some memory locations at the RAM nodes in the discriminator corresponding to the trained class are accessed according to each input pattern. Each access increments by one the value stored in the respective location. During the classification phase, every discriminator retrieves the information addressed by the input pattern. Each RAM node accessed this way outputs one (1) if the memory position in question holds a value higher than the bleaching threshold, and zero (0) otherwise. A discriminator response is the sum of the outputs of each of its RAM nodes, as seen in Fig. 1. In a WiSARD multi discriminator arrangement, the discriminator with the highest response is chosen for the classification, as depicted by Fig. 2. If two or more discriminators share the highest response then the bleaching threshold must be incremented by one and a new classification iteration is performed. Training and classification can be interleaved during runtime. By doing so, WiSARD can be employed in continuous (online) learning tasks.

2.1.2. Comparison with other learning models

WiSARD may resemble other learning models in some aspects, while being intrinsically different in others. Bayesian classifiers [25] also learn by counting occurrences of events regarding attributes values, but without explicit use of binary features. Curve