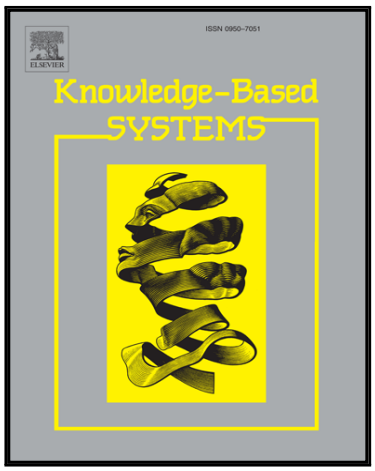

Classifiers consensus system approach for credit scoring

Maher Ala'raj , Maysam F. Abbod

PII: S0950-7051(16)30056-9
DOI: 10.1016/j.knosys.2016.04.013
Reference: KNOSYS 3483

To appear in: *Knowledge-Based Systems*

Received date: 18 August 2015
Revised date: 14 April 2016

Please cite this article as: Maher Ala'raj , Maysam F. Abbod , Classifiers consensus system approach for credit scoring, *Knowledge-Based Systems* (2016), doi: 10.1016/j.knosys.2016.04.013

Classifiers consensus system approach for credit scoring

Maher Ala'raj*, Maysam F. Abbod

Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, United Kingdom

ABSTRACT

Banks take great care when dealing with customer loans to avoid any improper decisions that can lead to loss of opportunity or financial losses. Regarding this, researchers have developed complex credit scoring models using statistical and artificial intelligence (AI) techniques to help banks and financial institutions to support their financial decisions. Various models, from easy to advanced approaches, have been developed in this domain. However, during the last few years there has been marked attention towards development of ensemble or multiple classifier systems, which have proved their ability to be more accurate than single classifier models. However, among the multiple classifier systems models developed in the literature, there has been little consideration given to: 1) combining classifiers of different algorithms (as most have focused on building classifiers of the same algorithm); or 2) exploring different classifier output combination techniques other than the traditional ones, such as majority voting and weighted average. In this paper, the aim is to present a new combination approach based on classifier consensus to combine multiple classifier systems (MCS) of different classification algorithms. Specifically, six of the main well-known base classifiers in this domain are used, namely, logistic regression (LR), neural networks (NN), support vector machines (SVM), random forests (RF), decision trees (DT) and naïve Bayes (NB). Two benchmark classifiers are considered as a reference point for comparison with the proposed method and the other classifiers. These are used in combination with LR, which is still considered the industry-standard model for credit scoring models, and multivariate adaptive regression splines (MARS), a widely adopted technique in credit scoring studies. The experimental results, analysis and statistical tests demonstrate the ability of the proposed combination method to improve prediction performance against all base classifiers, namely, LR, MARS and seven traditional combination methods, in terms of average accuracy, area under the curve (AUC), the H-measure and Brier score (BS). The model was validated over five real-world credit scoring datasets.

Keywords: credit scoring; consensus approach; multiple classifier systems; classifier ensembles; classification

Introduction

1.1. Background

Credit granting to lenders is considered a key business activity that generates profits for banks, financial institutions and shareholders, as well as contributing to the community. However, it can also be a great source of risk. The recent financial crises resulted in huge losses globally and, hence, increased the attention paid by banks and financial institutions to credit risk models. That is, as a result of the crises, banks are now cognisant of the need to adopt rigorous credit evaluation models in their systems when granting a loan to an individual client or a company. The problem associated with credit scoring is that of categorizing potential borrowers into either good or bad. Models are developed to help banks to decide whether to grant a loan to a new borrower or not using their data characteristics (Hand and Henley, 1997). The area of credit scoring has become a widely researched topic by scholars and the financial industry (Kumar and Ravi, 2007; Lin et al. 2012) since the seminal work of Altman in 1968 (Altman, 1968). Subsequently, many models were proposed and developed using statistical approaches, such as logistic regression (LR) and linear discriminate analysis (LDA) (Desai et al., 1996; Baesens et al., 2003). Recently, the Basel Committee on Banking Supervision (Lessmann et al., 2015) requested that all banks and financial institutions to have rigorous and complex credit scoring systems in order to help them improve their credit risk levels and capital allocation.

Despite developments in technology, LR is still the industry-standard baseline model used for building credit scoring models (Lessmann et al., 2015); many studies have demonstrated that artificial intelligence (AI) techniques, such as neural networks (NN), support vector machines (SVM), decision trees (DT), random forests (RF) and naïve Bayes (NB), can be substitutes for statistical approaches in building credit scoring models (Atiya, 2000; Van Gestel et al., 2003; Wang et al., 2012; Verikas et al., 2011; Hsieh and Hung, 2010; Zhou, 2013).

In practice, real historical datasets are used in order to develop credit-scoring models; these datasets might differ in size, nature, and the information or characteristics it holds, whilst individual classifiers might not be able to capture different relationships of these datasets characteristics. As a result, researchers have employed hybrid modelling

* *Corresponding author*. Tel.: +447466925096.
E-mail address: maher.ala'raj@brunel.ac.uk