Applied Soft Computing 62 (2018) 73–85

Contents lists available at ScienceDirect

Applied Soft Computing

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

Full Length Article

Predicting corporate investment/non-investment grade by using interval-valued fuzzy rule-based systems—A cross-region analysis

Petr Hajek

University of Pardubice, Institute of System Engineering and Informatics, Faculty of Economics and Administration, Studentská 84, Pardubice 532 10, Czech Republic

ARTICLE INFO

Article history:
Received 24 February 2017
Received in revised form 30 August 2017
Accepted 19 October 2017
Available online 27 October 2017

Keywords:
Interval-valued fuzzy rule-based systems
Evolutionary algorithms
Financial distress
Credit rating

ABSTRACT

Systems for predicting corporate rating have attracted considerable interest in soft computing research due to the requirements for both accuracy and interpretability. In addition, the high uncertainty associated primarily with linguistic uncertainties and disagreement among experts is another challenging problem. To overcome these problems, this study proposes a hybrid evolutionary interval-valued fuzzy rule-based system, namely IVTURS, combined with evolutionary feature selection component. This model is used to predict the investment/non-investment grades of companies from four regions, namely Emerging countries, the EU, the United States, and other developed countries. To evaluate prediction performance, a yield measure is used that combines the return and default rates of companies. Here, we show that using interval-valued fuzzy sets leads to higher accuracy, particularly with the growing granularity at the fuzzy partition level. The proposed prediction model is then compared with several state-of-the-art evolutionary fuzzy rule-based systems. The obtained results show that the proposed model is especially suitable for high-dimensional problems, without facing rule base interpretability issues. This finding indicates that the model is preferable for investors oriented toward developed markets such as the EU and the United States.

1. Introduction

Credit ratings are increasingly considered to be an important decision-making support to financial market participants such as investors, financial institutions and regulators. Credit ratings play an important role in the financial system by reducing information asymmetry between investors and borrowers. Credit ratings are assigned to issuers, as well as to specific debt issues, such as bonds, notes, and other debt securities. In issuer's credit ratings, an issuer's overall capacity and willingness to meet its financial obligations is addressed. This evaluation is based on a complex analysis performed by professionals who interpret both financial and non-financial information from multiple sources. Since this evaluation can be slow and costly, automatic credit rating prediction has become a central problem in artificial intelligence (AI) research [1]. Prediction models have been extensively developed to replicate and explain the credit rating processes performed by credit rating agencies [2].

A wide range of AI methods have been applied to predict credit ratings, including statistical classifiers [3], decision trees [1], neural networks (NNs) [4,5], support vector machines (SVMs) with both supervised [6,7] and semi-supervised learning [8], case-based reasoning [9], artificial immune systems [10], rough sets [11], fuzzy rule-based systems (FRBSs) [12], and ensemble approaches [10,13]. Recent efforts have also indicated that AI methods should be integrated into the feature selection process to improve prediction accuracy [1]. Finally, the paradigm of soft computing [14] has recently provided the most encouraging results in related problems such as financial failure prediction [15,16] and a consumer's credit scoring [17]. This refers to the integration of different, seemingly unrelated, AI methods such as FRBSs, NNs, evolutionary algorithms (EAs), rough set theory and probabilistic reasoning in various combinations to exploit their strengths. However, significantly insufficient attention has been paid to its application in corporate credit rating prediction.

Most importantly, the hybridization between FRBSs and EAs provides advantages that are desirable in imitating the credit rating process. First, evolutionary FRBSs provide good interpretability in terms of fuzzy if-then rules (in contrast to non-if-then fuzzy classifiers) and, thus, simulate the credit rating decision-making process of financial experts. Note that FRBSs, similarly as other possibilistic classifiers, may assign a soft class label with degrees of membership in each class [18]. This is similar to probabilis-

E-mail address: petr.hajek@upce.cz

https://doi.org/10.1016/j.asoc.2017.10.037
1568-4946/

tic classifiers that usually have the posterior probabilities for the classes as output. However, in the credit rating process, experts use linguistic labels to represent the partial truth of their opinions, rather than partial knowledge. Thus, the experts are able to verify the classification paradigm, for example the consistency and completeness of the rule base. Second, EAs are employed to learn or tune different components of FRBSs such as rule bases, the antecedents and consequents of if-then rules, parameters of membership functions (MFs), and so on. Thus, EAs enable the automatic design of FRBSs through their capability to encode and evolve rule antecedent aggregation operators, different rule semantics, rule base aggregation operators and defuzzification methods [19]. In addition, decision makers can fix some components of FRBSs in order to improve interpretability and accuracy. Owing to these qualities, evolutionary FRBSs represent one of the most popular approaches in the soft computing literature [20]. However, determining the precise values of MFs can be problematic in many application domains due to the uncertainties associated with dynamic unstructured environments, linguistic uncertainties, disagreement among experts, and noise in the data [21]. Therefore, several generalizations of FRBSs have been developed to design MFs effectively. Interval-valued FRBSs (IVFRBSs) [22] are widely considered to be the most important representatives of these generalizations. Recently, methods have been developed to optimize the design of IVFRBSs that, compared with traditional FRBSs, provide an additional degree of freedom and flexibility in handling uncertainty [23,24].

This study evaluates IVFRBSs optimized by EAs, namely IVTURS (interval-valued fuzzy rule-based system with tuning and rule selection) [25], to predict corporate investment/non-investment grade. First, the genetic feature selection process is carried out to obtain IVFRBSs with both comprehensible rule bases and high prediction accuracy. Second, we employ a range of state-of-the-art evolutionary FRBSs to compare their performance with IVFRBSs when predicting the grades of companies in four regions, namely the United States, the EU, other developed and Emerging countries. To measure prediction performance, we use a yield measure that is suitable for investors' decision making because it combines return and default rates. We also measure the interpretability of FRBSs at both the rule base and the fuzzy partition level.

The proposed system is aimed to provide an accurate and interpretable decision-support tool mainly to investors. Investors may use the system to match the relative credit risk of an issuer with their own risk tolerance in making investment decisions and portfolio management. However, the system may also be useful for other market participants, such as companies to make financing decisions (on the cost of capital and capital structure, for example) or financial institutions to assess counterparty risk.

As opposed to previous studies using AI to predict corporate rating grades, the proposed methodology integrates: (1) interval-valued fuzzy sets to address the issue of high linguistic uncertainty of expert opinions; (2) genetic feature selection to achieve high interpretability and verifiability of the rule base; and (3) evolutionary optimization of FRBSs to guarantee a high accuracy of the prediction system. In addition, this is, to our best knowledge, the first study comparing the performance of rating prediction models across multiple regions.The remainder of this paper has been organized in the following way. Section 2 provides a brief overview of soft computing applications in financial distress prediction. Section 3 describes the research methodology, including missing data treatment and the feature selection process. Section 3 also introduces the design of IVFRBSs for corporate investment/non-investment grade prediction. Section 4 presents the datasets used in this study. Section 5 examines the performance of IVFRBSs, mainly in terms of the yield obtained. Section 6 concludes and discusses both the results and the possible future research directions.

2. Soft computing in corporate financial distress prediction

Financial distress can be defined as a situation that clearly shows an enterprise's financial difficulty, such as statutory bankruptcy and credit default [26]. Credit default is usually estimated by rating grades. In the domain of financial distress prediction using soft computing, the research to date has tended to focus on bankruptcy rather than rating grades prediction. Here, we review a large and growing body of the literature that has investigated various combinations of AI methods in financial distress prediction.

Several papers have systematically reviewed recent research in this field. Kumar and Ravi [27] conducted a comprehensive review of the work on bankruptcy prediction during the 1968–2005 period. Based on this overview, the authors claimed that the most successful prediction models, rather than using a single method, are based on hybrid soft computing systems. These were categorized into (1) ensemble classifiers [28], (2) intelligent feature selection combined with classification [29], and (3) tightly integrated hybrid systems (evolutionary NNs, fuzzy NNs, etc.) [30]. However, later review studies have regarded only the latter two as soft computing approaches [31]. In their review, Verikas et al. [31] identified the following soft computing approaches for financial distress prediction: (1) genetic algorithms (GAs) in hybrid techniques (to select a subset of input features, to find the appropriate hyper-parameter values of a predictor, or to determine predictor parameters) [32,33]; (2) rough sets in hybrid techniques (to select a subset of input features) [34]; (3) fuzzy set theory-based techniques (to increase transparency) [12]; (4) self-organizing maps in hybrid systems (for data exploration and visualization) [35]; and (5) combining traditional and soft computing techniques [36]. Lin et al. [37] categorize four soft computing approaches, GAs, Group method of data handling, rough sets, and fuzzy sets.

Sun et al. [26] also categorized soft computing approaches. In the first type, one algorithm (usually GAs or rough sets) is applied to choose the features of another classification algorithm. Second, one algorithm is applied to optimize the parameters for another classification algorithm. Finally, a new classification algorithm is produced by integrating two or more algorithms. For example, Cheng et al. [38] embedded logit analysis into the output layer of radial basis function NNs, whereas Chaudhuri and De [39] developed fuzzy SVMs to handle uncertainty and impreciseness in corporate data.

Other hybrid systems combine multiple criteria decision-making methods with soft computing methods. Wu and Hsu [40] employed TOPSIS to determine the optimal classifier and subsequently extracted knowledge from the classifier by using decision trees. Shen and Tzeng [41] combined feature selection using rough sets with multiple criteria decision-making methods to collect the knowledge of domain experts.

Taken together, previous studies of financial distress prediction have reported that (1) the feature selection process improves prediction accuracy, and (2) hybrid systems improve both accuracy and transparency. Although extensive research has been carried out on financial distress prediction, to our best knowledge, no single study exists that adequately covers the advantages of evolutionary IVFRBSs in a financial distress prediction model. However, evolutionary IVFRBSs have recently been employed in related financial applications, namely credit scoring, fraud detection and stock market trend prediction [42].

3. Research methodology

3.1. Rating grades and datasets

A credit rating (rating grade) is an expert evaluation of the general creditworthiness of an obligor. This evaluation is conducted