Knowledge-Based Systems 39 (2013) 224–239

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry

You-Shyang Chen [a,*], Ching-Hsue Cheng [b]

[a] Department of Information Management, Hwa Hsia Institute of Technology, 111, Gong Jhuan Rd., Chung Ho District, New Taipei City 235, Taiwan
[b] Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan

ARTICLE INFO

Article history:
Received 28 March 2012
Received in revised form 12 November 2012
Accepted 13 November 2012
Available online 28 November 2012

Keywords:
Credit rating
Factor analysis
Attribute reduction
Rough Set Theory (RST)
Minimum Entropy Principle Approach (MEPA)

ABSTRACT

Banks are important to national, and even global, economic stability. Banking panics that follow bank insolvency or bankruptcy, especially of large banks, can severely jeopardize economic stability. Therefore, issuers and investors urgently need a credit rating indicator to help identify the financial status and operational competence of banks. A credit rating provides financial entities with an assessment of credit worthiness, investment risk, and default probability. Although numerous models have been proposed to solve credit rating problems, they have the following drawbacks: (1) lack of explanatory power; (2) reliance on the restrictive assumptions of statistical techniques; and (3) numerous variables, which result in multiple dimensions and complex data. To overcome these shortcomings, this work applies two hybrid models that solve the practical problems in credit rating classification. For model verification, this work uses an experimental dataset collected from the Bankscope database for the period 1998–2007. Experimental results demonstrate that the proposed hybrid models for credit rating classification outperform the listing models in this work. A set of decision rules for classifying credit ratings is extracted. Finally, study findings and managerial implications are provided for academics and practitioners.

1. Introduction

The world financial crisis during 2007–2009 was due to reckless and unsustainable lending practices under deregulation and securitization of United States (US) mortgages, which were marketed as investments to global individual investors and financial institutions. The risks of a broad-based credit boom arose from a near-global speculative bubble in real estate and equities, eventually exposing other risky loans and inflated asset values, and instigating a global recession. Clearly, the emergence of subprime loan losses from 2007 to present rapidly reduced economic activity and saw numerous financial institutions and firms facing extreme financial difficulties, distress, or even bankruptcy. Despite a significant easing of national fiscal and monetary policy in US in an effort to stem the global recession, the world has failed to shake the financial crisis. Global investors continue to face difficult challenges, particularly in relation to banks in Europe, particularly in Greece, Italy, Portugal, and Spain, which currently have the same debt problems [29]. Banks typically face numerous risks, including credit, debt, interest rate, currency, liquidity, and systematic risk. Serial bank runs and other blows to financial sector confidence inevitably result from extreme events, such as the current financial crisis, which has seriously jeopardized regional and national economic stability [48]. Banks are clearly important to national, and even global, economic stability. However, relative to other industries, banking stability is very dependent on trust and reputation, which is particularly true for large banks. Global banking investors should therefore protect their profits by identifying high-quality targets. Consequently, an indicator is urgently needed to identify a bank's financial status and operational competence.

Market investors have used numerous indicators to identify superior investment targets in seeking increased profits. Credit ratings [9,18] evaluate the attractiveness of banks as investments. When properly assigned by rating agencies, such as Standard and Poor's (S&P), Moody's, and Fitch, they are invaluable to financial market participants, providing objective opinions about credit worthiness, investment risk, and default probability. Interested parties include owners, customers, management, personnel, investors, competitors, suppliers, creditors, media, regulatory agencies, researchers, and special-interest groups. Each group uses credit ratings in its own way [47]. For example, credit ratings are extremely important to stock market investors. Although the process of assigning a credit rating requires an enormous amount of time and resources [24], classification models based on financial ratios [13], such as capital adequacy, asset quality, management competence, liquidity risk, sensitivity to market risk (CAMELS) [6] and Earnings Before Interest and Taxes (EBITs), can simplify this

* Corresponding author. Tel.: +886 2 8941 5143; fax: +886 2 8941 5142.
E-mail address: ys_chen@cc.hwh.edu.tw (Y.-S. Chen).

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.11.004

process [44]. Financial ratios are typically employed to evaluate bank financial and operational competence, and rate overall management effectiveness based on quarterly and/or annual sales and investment performance. Financial ratios are widely used for modeling by both practitioners and researchers, and have been expressed in various forms [25,56]. Problems associated with assigning credit ratings resemble those related to forecasting financial crises and bankruptcy [13,30,44,53], which can be developed to construct early warning systems (EWSs) [6] using classification models based on financial ratios.

Since the 1960s, numerous studies have constructed models for predicting financial crises and bankruptcy. These studies have applied both statistical methods and artificial intelligence (AI) techniques, including multiple discriminant analysis (MDA) [40], a logistic model [34], support vector machines (SVMs) [9,24,30], and neural networks (NNs) [33]. Although these statistical methods are simple and their outcomes are easy to explain, their explanatory power is inferior to that of AI techniques. This creates decision-making difficulties, as policy-makers cannot fully comprehend and follow the results of the models they use. Moreover, almost all studies comparing the efficiency of these methods found that performance was highly dependent on the application field [8], study goals [17], context and data [27], or user experience [3]. Thus, we recommend employing AI techniques to develop efficient classifiers for forecasting. Artificial intelligence techniques, which have been extensively used when generating credit ratings, have outperformed statistical methods [9,24]. Particularly, intelligent hybrid systems integrate several models for processing classification problems [2,50,55]. In practice, an ensemble classifier outperforms stand-alone models [43,44]. Given the limitations of statistical methods and AI techniques in stand-alone models, an intelligent hybrid model is needed that maximizes the advantages of statistical methods and AI techniques while minimizing their limitations. Interest in designing and applying various intelligent hybrid models has increased considerably over the last decade [43]. To improve prediction performance and increase investor profits, a reliable forecasting tool based on a hybrid model is required for classifying bank credit ratings.

Designing a rule-based model that can reasonably and powerfully explain data is a significant trend in knowledge discovery. Notably, AI techniques for classification can automatically extract knowledge-based decision rules from a dataset and construct different model representations to explain that dataset [54]. Market investors are very interested in rule-based models that are based on AI techniques and germane to the global banking industry. Research to improve models that solve credit rating problems is valuable for two reasons. First, although the financial industry focuses on investors seeking personal benefit, AI techniques have rarely been used in credit rating research to generate comprehensive decision rules, particularly when compared with statistical methods; therefore, this work fills this knowledge gap. Second, the authors have extensive experience in the financial industry, about 14 years in total, and thus have relevant knowledge.

Because each interested party has an opinion of how to best apply intelligent hybrid systems to real-world problems, particularly the current financial crisis, a reliable model that predicts credit ratings is welcomed. To objectively address the practical problems of classifying bank credit ratings and generate decision rules in the form of knowledge-based systems, this work applies two hybrid models. This work has the following four objectives: (1) implement the two hybrid models that use rough sets (RSs) to classify credit ratings in the global banking industry, minimize the number of selected attributes and generated rules, and increase prediction accuracy; (2) examine the main determinants influencing credit ratings; (3) assess the performances of the proposed hybrid models; and (4) generate comprehensive decision rules that can be applied to knowledge-based systems using RSs and the LEM2 algorithm, and provide reasonable explanatory power to interested parties.

The remainder of this paper is organized as follows. Section 2 reviews the literature on credit rating classification. Section 3 describes the proposed hybrid models and algorithms. Section 4 gives verification details and compares the models. Finally, Section 5 draws conclusions and provides directions for future research.

2. Literature review

This section reviews credit rating literature, including that associated with rough set theory (RST), the LEM2 algorithm for rule extraction, rule filter (RF), minimize entropy principle approach (MEPA), and factor analysis (FA).

2.1. Credit rating

Credit ratings are assessments of the creditworthiness of issuers or issues, and involve a hierarchical ranking process by which credit is classified into different risk categories [35] by credit rating agencies. Credit rating agencies, such as S&P, Moody's, and Fitch, assess the capacity of entities to fulfill their financial commitments. A credit ratings is a benchmark measure of default probability, namely, of a debtor failing to meet its obligations under the debt contract, and of the expected associated losses. Low ratings indicate high risk of default. Investors use credit ratings to indicate the likelihood of receiving their money back in accordance with the investment terms [61].

Numerous statistical solutions, including multiple linear regression (MLR), MDA, logit, and multi-criteria decision aid (MCDA) models, have been applied to solve credit rating classification problems. Horrigan [23] first applied MLR to forecast corporate bond ratings. The dataset included 150 training and 150 testing data from Moody's and S&P. Six financial attributes and nine classes were utilized; the accuracy rates for Moody's and S&P were 58% and 52%, respectively. Pinches and Mingo [40] applied FA to determine financial attributes, and then applied MDA to construct linear functions. The training data comprised 180 samples of industrial bonds and the testing data comprised 96 samples from Moody's. The data included six conditional attributes and five classes; the achieved accuracy was 65%. Kaplan and Urwitz [26] employed and ordered a logit model to solve order problems involving credit ratings. Their study gathered 120 training samples involving industrial bonds and 53 testing samples from Moody's. The model had 15 financial attributes and four classes, and had an accuracy of 65%. Pasiouras et al. [36] designed a MCDA model that applied a Multi-group Hierarchical Discrimination (MHDIS) approach to rate the credit of Asian banks. Their study selected 153 commercial banks from Fitch and ten attributes, and grouped ratings into five classes. Analytical results showed that the MHDIS model was more accurate (66.03%) than discriminant analysis (53.73%) and ordered logistic regression (47.55%).

Due to rapid development of information technology (IT), AI techniques for classification, such as NNs and SVMs, have also been employed to solve credit rating classification problems. Dutta and Shekhar [15] first applied NNs to rate bonds. The dataset used 47 samples and included six and ten financial attributes and two classes to predict bond ratings, achieving an accuracy of 83.3%. Kim et al. [28] compared AI techniques, including the NN model and rule-based expert system ID3, with statistical techniques, including linear regression, discriminant analysis, and logistic analysis, for bond rating. The dataset obtained from S&P comprised 228 samples, eight financial attributes, and six classes. The NN model had the best accuracy (55.17%), followed by logistic analysis