

Chapter 82

Default Probability Prediction of Credit Applicants Using a New Fuzzy KNN Method With Optimal Weights

Abbas Keramati

University of Tehran, Iran

Niloofar Yousefi

University of Central Florida, USA

Amin Omidvar

Amirkabir University of Technology, Iran

ABSTRACT

Credit scoring has become a very important issue due to the recent growth of the credit industry. As the first objective, this chapter provides an academic database of literature between and proposes a classification scheme to classify the articles. The second objective of this chapter is to suggest the employing of the Optimally Weighted Fuzzy K-Nearest Neighbor (OWFKNN) algorithm for credit scoring. To show the performance of this method, two real world datasets from UCI database are used. In classification task, the empirical results demonstrate that the OWFKNN outperforms the conventional KNN and fuzzy KNN methods and also other methods. In the predictive accuracy of probability of default, the OWFKNN also show the best performance among the other methods. The results in this chapter suggest that the OWFKNN approach is mostly effective in estimating default probabilities and is a promising method to the fields of classification.

1. INTRODUCTION

The increased demand for consumer credit has led to an intense competition in credit industry. Therefore, credit managers have to develop and apply machine learning methods to handle analyzing credit data in order to save time and reduce errors. Credit scoring can be defined as a technique that helps lenders decide whether to grant credit to the applicants with respect to the applicants' characteristics such as age,

DOI: 10.4018/978-1-5225-5643-5.ch082

income and marital status (M. C. Chen & Huang, 2003). In recent years, several quantitative methods have been proposed for credit risk evaluation. Among all existent approaches, data mining methods have found more popularity than the others because of their ability in discovering practical knowledge from the database and transforming them into useful information. The first researches on credit scoring were done by (Fisher, 1936) and (Durand, 1941) who applied linear and quadratic discriminant analysis respectively to categorize credit applications as “good” or “bad” ones.

Credit scoring models are known as statistical models which have been widely used to predict the default risk of individuals or companies. These are multivariate models which use the main economic and financial indicators of a company or the individuals’ characteristics such as age, income and marital status as input and assign them a weight which reflects their relative importance in predicting the default. The result is an index of creditworthiness that is expressed as a numerical score measuring the borrower’s probability of default (PD). Default probability is the likelihood that a borrower will be unable to repay a loan and fall into default.

The initial credit scoring models were devised in the 1930s by authors such as (Fisher, 1936) and (Durand, 1941) and have been developed by the studies of (Beaver, 1967), (Altman, 1968) and others from 1960. Many credit scoring methods have been proposed so far, but many of them only focus on classifying customers instead of scoring them. From the viewpoint of risk management, assigning a score to each borrower will be more meaningful than classifying him as a “good” or “bad” applicant. The most important issue in credit scoring is to assign a specific PD to each customer according to his characteristics and financial/economic indicators. In this article we will present a method which is capable of classifying and scoring applicants simultaneously.

In this study, we present a new method which is capable of estimating a default probability for each borrower in addition to classifying him. It will be shown that the PDs generated from this method are very close to their real values. The other advantage of the proposed model is that it does not need any a priori information about the structure of the data. The conventional methods for estimating PD such as discriminant analysis (DA) or logistic regression (LR) are based on fairly unrealistic assumptions (the normality of the data for DA and the existence of a logistic relationship between input variables and PD for LR).

In the next section, we review the most widely used data mining methods and the articles associated with each data mining method applied for credit scoring task from 2000 to 2010. Section 3 describes our methodology for doing this research in detail. Empirical results generated by Optimally Weighted Fuzzy K-Nearest Neighbor (OWFKNN) method on two real world data sets are shown in section 4. Section 5 contains some concluding remarks.

2. LITERATURE SURVEY

This paper presents a comprehensive review of literature related to application of data mining techniques in credit scoring published in academic journals between 2000 and 2010.

For doing this research, we searched online journal databases which some of them are referenced here. These databases include Science Direct; Scopus, Emerald, Springer, IEEE Explore, Jstore, John Wiley, Academic Search Premier. The searched Journals include Expert Systems with Applications, European Journal of Operational Research, Computers & Operations Research, Computer Engineering and Ap-

35 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/default-probability-prediction-of-credit-applicants-using-a-new-fuzzy-knn-method-with-optimal-weights/205862

Related Content

Integration of Management of Quality of Web Services in Service Oriented Architecture

M. A. Serhani, M. E. Badidi, A. Benharref, R. Dssouliand H. Sahraoui (2008). *Intelligent Information Technologies and Applications* (pp. 190-220).

www.irma-international.org/chapter/integration-management-quality-web-services/24266

Optimizing Software Development Cost Estimates Using Multi-Objective Particle Swarm Optimization

Tad Gonsalves, Kei Yamagishi, Ryo Kawabataand Kiyoshi Itoh (2010). *Artificial Intelligence Applications for Improved Software Engineering Development: New Prospects* (pp. 46-65).

www.irma-international.org/chapter/optimizing-software-development-cost-estimates/36441

Fuzzy Rough Support Vector Machine for Data Classification

Arindam Chaudhuri (2016). *International Journal of Fuzzy System Applications* (pp. 26-53).

www.irma-international.org/article/fuzzy-rough-support-vector-machine-for-data-classification/151534

Supporting Structured Group Decision Making through System-Directed User Guidance: An Experimental Study

Harold J. Lagroue III (2010). *Methodological Advancements in Intelligent Information Technologies: Evolutionary Trends* (pp. 261-278).

www.irma-international.org/chapter/supporting-structured-group-decision-making/38530

Hybrid System based on Rough Sets and Genetic Algorithms for Medical Data Classifications

Hanaa Ismail Elshazly, Ahmad Taher Azar, Aboul Ella Hassanienand Abeer Mohamed Elkorany (2013). *International Journal of Fuzzy System Applications* (pp. 31-46).

www.irma-international.org/article/hybrid-system-based-rough-sets/101768