

Rough sets and logistic regression analysis for loan payment

Bahaddin Ruzgar, Nursel Selver Ruzgar

Abstract—Risk classification is an important part of the financial processes. In small business loans, there is always a risk for nonpayment or non-refunding of loans though very detailed examinations are made about the company. In this study, behaviors that increase the risk in loans or causing non-refunding are tried to be determined by using the rough Set (RS) approach and logistic regression (LR). For this purpose, 121 regularly refunded and 121 irregularly refunded loans, drawn from a bank in 2006 year, were randomly selected and their attributes were examined in 2007. Examination is made in three sections for qualitative variables, for quantitative variables and for both qualitative and quantitative variables. As a result, RS model is applicable to a wide range of practical problems pertaining to loan payment prediction, but LR does not classify refund or non-refund of loan payment as good as RS, so LR can not be used for prediction. Moreover, the results show that the RS model is a promising alternative to the conventional methods for financial prediction. In fact, RS gives the attributes that affect the results negatively or positively with their measures which are used for predictions.

Keywords—Classification, loan payment, logistic regression, rough sets.

I. INTRODUCTION

The financial system plays a crucial role in economic development as responsible for the allocation of resources over time and among different alternatives of investment by pricing the postponement of consumption and pricing the risk [1]. Decompositions of balances on financial systems affect markets badly. Therefore, these may force sound banks to go to bankrupt, large or small business to go to failure, or loans payments to go to failure. In this paper, we propose an approach to predict small business loan payment failure based on genetic algorithm techniques of RS and LR [2].

Business failure prediction is a scientific field which many academic and professional people have been working for, at least, the three last decades. The high individual and social costs encountered in corporate bankruptcies make this decision problem very important to parties such as auditors, management, government policy makers, and investors. Also,

financial organizations, such as banks, credit institutions, clients, etc., need these predictions for firms in which they have an interest (of any kind) [3]. Many financial decisions involve the classification of a set of observations (firms, stocks) into one or several groups of categories, what leads many researches to apply operational research methods to management problems. Methods such as discriminant analysis (DA), logit analysis, RS, recursive partitioning algorithm, and several others have been used in the past for the prediction of business failure. Although some of these methods led to models with a satisfactory ability to discriminate between healthy and potentially risky (candidates for bankruptcy) businesses, they suffer from some limitations, often due to the unrealistic assumption of statistical hypotheses or due to a confusing language of communication with the decision makers [4]. There exists an extensive literature devoted to the study of classification problems in the financial field, i.e. the work by Chen and Yeh [5] dealing with financial management or the works by Dimitras et al. [6] and Tam [7] in business failure [2]. Over the last 35 years, academic researchers from all over the world have dedicated their time to the search for the best corporate failure prediction model that can classify companies according to their financial health or failure risk [6], [8], [9]. Most papers deal with insurance audits, purchase intentions, purchase channel studies, methodologies for investigating customer purchasing intentions, and customer satisfaction [10]. Many research papers have quantified the problem in order to simplify the parameters, such as social parameters, and use statistical tools to analyze data. This approach, however, is only good for crisp types of data sets and certain data values. If the value of data is continuous or uncertain we must apply fuzzy theory [11]. In this paper, two methods, RS approach and LR are used to provide a set of rules able to discriminate between healthy and failing firms in order to predict business loan payments [3].

The Rough Set Theory (RST) has been successfully applied in many real-life problems in medicine, pharmacology, engineering, banking, financial and market analysis and others. The RS methodology has found many real-life applications [12]. RST overlaps with many other theories, such as fuzzy sets, evidence theory, and statistics. Nevertheless, it can be viewed in its own right as an independent, complementary, and no competing discipline [13]. The most important problems that can be solved by RST are: finding description of sets of objects in terms of attribute values, checking dependencies (full or partial) between attributes, reducing attributes, analyzing the significance of attributes, and generating decision rules [14], [15]. The application of the RS approach in

Manuscript received December 14, 2007; Revised version received March 8, 2008.

N. S. Ruzgar, is with Vocational School of Social Science, Marmara University, Istanbul, Turkey, (phone: 90-212-517 20 16; fax: 90-216-517 20 12; e-mail: nrugzar@marmara.edu.tr).

B. Ruzgar is with the Actuaries Department, Banking and Insurance School, Istanbul, Marmara University, Turkey, (phone: 90-216-414 99 89; fax: 90-216-347 50 86; e-mail: bruzgar@marmara.edu.tr).