A Comparative Study of Statistical and Rough Computing Models in Predictive Data Analysis

Debi Acharjya, School of Computing Science and Engineering, VIT University, Vellore, India
A. Anitha, School of Information Technology and Engineering, VIT University, Vellore, India

ABSTRACT

Information and technology revolution has brought a radical change in the way data are collected. The data collected is of no use unless some useful information is derived from it. Therefore, it is essential to think of some predictive analysis for analyzing data and to get meaningful information. Much research has been carried out in the direction of predictive data analysis starting from statistical techniques to intelligent computing techniques and further to hybridize computing techniques. The prime objective of this paper is to make a comparative analysis between statistical, rough computing, and hybridized techniques. The comparative analysis is carried out over financial bankruptcy data set of Greek industrial bank ETEVA. It is concluded that rough computing techniques provide better accuracy 88.2% as compared to statistical techniques whereas hybridized computing techniques provides still better accuracy 94.1% as compared to rough computing techniques.

KEYWORDS

Almost Indiscernibility, Correlation, Equivalence Class, Fuzzy Proximity Relation, Fuzzy Relation, Mean Percentile Error, Mean Square Error, Neural Network, Prediction, Regression Analysis, Rough Set

1. INTRODUCTION

Few decades before, computer was a simple device used for doing computations, and calculations in a limited area. But emergence of networking and communication technologies, has replaced the role of computer from stand alone system to distributed systems. Simultaneously, the processing speed is considerably increased. This helps in processing data at a greater speed. At present age, enormous amount of data are exchanged, generated, stored, and manipulated through the internet and through numerous sources. But, what is the need of such huge accumulated data unless we extract or predict some useful information from it. So data analysis, information retrieval, and prediction of decisions for unseen associations is of recent research. Additionally, the branch of data mining concerned about the prediction of future probabilities and trends are referred as predictive analysis. It deals with the variables that can be measured based on other single or multiple factors to predict the decision.
In traditional approach of predictive modeling, data are collected, a statistical model is formulated and predictions are made with validating the available data. But statistical methods have its own limitations and cannot produce better prediction when the data contains uncertainty. Further to handle uncertainty in predictive data analysis, many intelligent techniques such as rough set (Pawlak, 1982), rough set on fuzzy approximation space (De, 1999), rough set on intuitionistic fuzzy approximation space (Acharjya & Tripathy, 2009), and hybridization of these concepts with other techniques such as neural network, genetic algorithm, formal concept analysis (Tripathy, Acharjya & Ezhilarasi, 2011), etc., were developed. Moreover, predictive analysis is applied in numerous areas such as health sector, telecommunications, financial services, marketing, actuarial science, travel, pharmaceuticals etc. Our basic objective in writing this paper is to make a comparative study between statistical approach and some of the computational intelligent approach. To show the viability of comparison, financial bankruptcy dataset is used to measure the financial distress of the public firms.

Corporate bankruptcy plays a significant role in the field of finance, for the economic phenomena of a country. Policy makers, investors, managers, consumers, industry shares holders, are the prominent entities for the healthy and successful business world (Cielen, Peters, & Vanhoof, 2004). Business failure is a world-wide problem. To enhance the growth throughout the country, some mechanism should be available to predict the number of firms that may fail due to bankruptcy. Simultaneously, the failure serves as an index for the continuous development and robustness of a country’s economy (Min & Jeong, 2009; Zhang & Wu, 2011). The consequences raised by the corporate bankruptcies urge the researchers to carry out research work in this direction. Bankruptcy prediction technique is a vast area of finance and accounting research. The research on developing such prediction models initializes its process by focusing on various classification models to distinguish failed and non-failed firms. Such models are of major importance for the budgetary decision makers, as they serve as early-warning system for the failure probability of a corporate entity. To this end, varied traditional statistical methods are employed for predicting financial distress. As stated earlier, in this paper, we compare the statistical approach with various rough computing techniques, using the data collected from the Greek industrial bank, ETEVA, which finances industrial and commercial firms in Greece (Slowinski & Zopoundis, 1995; Greco, Matarazzo, Pappalardo, & Slowinski, 2005). Furthermore, it will help on forming a economic distress prediction system to provide information to the investors, policy makers, and monitoring organizations.

The rest of the paper is organized as follows: Section 2 provides literature review on the bankruptcy prediction models. Section 3, discuss the foundations of the techniques used for predictive data analysis whereas Section 4 explains about the data organization and the proposed research model using various predictive data analysis techniques followed by Section 5, that depicts an experimental comparative analysis, and the paper is concluded by conclusion in Section 6.

2. LITERATURE REVIEW

Bankruptcy prediction models can be classified into two broad categories: statistical and computational intelligence models. Since 1932, bankruptcy predictions are subject to formal analysis by the world (Fitzpatrick, 1932). The analysis was done based on the financial ratios but not with any statistical methods. Later in 1967, William beaver applied t-test for the evaluation of the financial ratios with a linear variable (Beaver, 1966). Instead of using a single variable, Edward I. Altman used multiple discriminant analysis with the pair-matched samples, along with various other stochastic models such as conditional logit model and probit models (Altman, 1968). However, the practical application of these statistical models are limited by their inherent strict assumptions such as linearity, normality,
Bio-Inspired Algorithms in Bioinformatics II
[www.igi-global.com/chapter/bio-inspired-algorithms-bioinformatics/10255?camid=4v1a](www.igi-global.com/chapter/bio-inspired-algorithms-bioinformatics/10255?camid=4v1a)

Creating Sound Glyph Database for Video Subtitling
[www.igi-global.com/chapter/creating-sound-glyph-database-for-video-subtitling/173434?camid=4v1a](www.igi-global.com/chapter/creating-sound-glyph-database-for-video-subtitling/173434?camid=4v1a)