

Chapter 28

Credit Scoring: A Constrained Optimization Framework With Hybrid Evolutionary Feature Selection

Pantelis Z. Lappas

Department of Statistics and Stochastic Modelling and Applications Laboratory, Athens University of Economics and Business, Greece

Athanasios N. Yannacopoulos

Department of Statistics, Athens University of Economics and Business, Greece

ABSTRACT

The main objective of this chapter is to propose a hybrid evolutionary feature selection approach for solving credit scoring problems subject to constraints. A hybrid scheme combining filter and wrapper-based approaches is proposed to develop an accurate credit scoring model with a high predictive performance. Initially, the minimum redundancy maximum relevance algorithm is applied to find an optimal set of features that is mutually and maximally dissimilar and can represent the response variable effectively, allowing for an ordering of features by their importance. Subsequently, an iterative procedure, where supervised machine learning algorithms such as the logistic regression and the linear-discriminant analysis are combined with an evolutionary optimization algorithm like the genetic algorithm, is applied to choose the feature subset that maximizes an appropriate classification measure according to the predefined features and subject to the predefined constraints. The performance of the proposed method is illustrated using standard credit scoring datasets.

INTRODUCTION

One of the most important developments in the field of risk management is financial risk management (Verbano & Venturini, 2011). An important type of financial risk is credit risk which is defined as the probability that a customer will default on a loan; namely, the potential failure of meeting his/her loan obligations (Thomas et al., 2002). Credit scoring is an important topic in the credit risk management and has been the major focus of the banking industry, especially in the aftermath of 2008 crisis.

DOI: 10.4018/978-1-7998-5077-9.ch028

Credit Scoring

Usually, data collected by banks include a large amount of features. Some of them may be inconsistent or redundant due to their high inter-correlation (Liu & Schumann, 2005). Therefore, a considerable effort in cleansing or pre-selection is needed to make the collected data useful for credit scoring. Choosing the feature subset that maximizes a chosen classification measure is an essential and crucial step of a typical credit scoring process. Data mining techniques such as feature selection are used due to the abundance of data and the imperative to discover knowledge from raw data (Liu & Motoda, 2012). Feature selection can be viewed as a dimensionality reduction technique where the number of features in a specific dataset is reduced to the minimum, while the performance of the data mining classification task is maximized (Wang et al., 2015).

The nowadays multi-dimensional nature of a customer's characteristics, the subjective judgment of a financial analyst (ad hoc decisions based on experience), as well as the complexity that occurs from designing with a high predictive performance necessitate the interaction of credit scoring with scientific areas such as statistics, machine learning, operations research and decision making.

The main objective of this chapter is to propose an accurate two-phase credit scoring model with a high predictive performance. Particularly, data mining feature selection techniques, supervised machine learning algorithms and an evolutionary optimization algorithm are used in hybrid synthesis to develop an optimal credit scoring model. In the first phase, the Minimum Redundancy Maximum Relevance (MRMR) algorithm is applied to find an optimal set of features that is mutually and maximally dissimilar and can represent the response variable effectively (Ding & Peng, 2005). Therefore, features can be ordered by their feature importance. Based on that, a financial analyst is allowed, according to his/her experience, to select an initial subset of features and define constraints (e.g., pairs of features of equal importance) so as to proceed to the second phase of the solution approach. The second phase is associated with an iterative procedure where supervised machine learning algorithms such as Logistic Regression (LR) and Linear Discriminant Analysis (LDA) are combined with an evolutionary optimization algorithm such as the Genetic Algorithm (GA) to choose the feature subset that maximizes a chosen classification measure according to the predefined features and subject to the predefined constraints.

The remainder of the chapter is organized as follows: Section 2 provides a brief of literature review of credit scoring problem and solution approaches applied to feature selection problems. A mathematical formulation of the feature selection problem, as well as some basic concepts regarding the MRMR, LR, LDA and GA are presented in Section 3. Section 4 introduces the proposed hybrid scheme based on supervised learning and evolutionary learning. In section 5, the authors illustrate the performance of the proposed scheme in terms of numerical experiments using standard credit scoring datasets. Section 6 concludes the chapter and provides directions for future work.

BACKGROUND

Taking into account the main risk management standards that have been developed (IRM Standard, ISO 31000, British Standard BS 31100 and COSO ERM Framework) a typical risk management process can be presented as a continuous cycle of the following five activities: (a) risk identification, (b) risk assessment that consists of risk analysis and risk evaluation, (c) risk treatment that contains risk strategies such as tolerate (accept/retain), treat (control/reduce), transfer (insurance/contract) and terminate (avoid/eliminate), (d) risk reporting and (e) risk review (Hopkin, 2010). In particular, credit scoring reflects a credit risk assessment activity that attempts to integrate two different functions within the credit risk

24 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/credit-scoring/261957

Related Content

Mining Competitors and Finding Winning Plans Using Feature Scoring and Ranking-Based CMiner++ Algorithm: Finding Top-K Competitors

Sujatha T., Wilfred Blessing N. R. and Suresh Palarimath (2023). *International Journal of Intelligent Information Technologies* (pp. 1-11).

www.irma-international.org/article/mining-competitors-and-finding-winning-plans-using-feature-scoring-and-ranking-based-cminer-algorithm/318670

Integrating Systems Modelling and Data Science: The Joint Future of Simulation and 'Big Data' Science

Erik Pruyt (2017). *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 822-840).

www.irma-international.org/chapter/integrating-systems-modelling-and-data-science/173361

A Systematic Literature Review on the Human-AI Partnership Roles in Higher Education

Minh Tung Tran (2024). *Enhancing and Predicting Digital Consumer Behavior with AI* (pp. 28-38).

www.irma-international.org/chapter/a-systematic-literature-review-on-the-human-ai-partnership-roles-in-higher-education/347194

Cyber Threat Intelligence for Industrial Automation: AI-Powered Strategies

Abuh Ibrahim Sani (2025). *AI-Enhanced Cybersecurity for Industrial Automation* (pp. 131-148).

www.irma-international.org/chapter/cyber-threat-intelligence-for-industrial-automation/379623

Extending a Conceptual Modeling Language For Adaptive Web Applications

Raoudha B. Djemaa, Ikram Amousand Abdelmajid B. Hamadou (2008). *International Journal of Intelligent Information Technologies* (pp. 37-56).

www.irma-international.org/article/extending-conceptual-modeling-language-adaptive/2434