

Chapter 25

Incorporating Qualitative Information for Credit Risk Assessment through Frequent Subtree Mining for XML

Novita Ikasari

Curtin University, Australia & University of Indonesia, Indonesia

Fedja Hadzic

Curtin University, Australia

Tharam S. Dillon

Curtin University, Australia

ABSTRACT

Credit risk assessment has been one of the most appealing topics in banking and finance studies, attracting both scholars' and practitioners' attention for some time. Following the success of the Grameen Bank, works on credit risk, in particular for Small Medium Enterprises (SMEs), have become essential. The distinctive character of SMEs requires a method that takes into account quantitative and qualitative information for loan granting decision purposes. In this chapter, we first provide a survey of existing credit risk assessment methods, which shows a current gap in the existing research in regards to taking qualitative information into account during the data mining process. To address this shortcoming, we propose a framework that utilizes an XML-based template to capture both qualitative and quantitative information in this domain. By representing this information in a domain-oriented way, the potential knowledge that can be discovered for evidence-based decision support will be maximized. An XML document can be effectively represented as a rooted ordered labelled tree and a number of tree mining methods exist that enable the efficient discovery of associations among tree-structured data objects, taking both the content and structure into account. The guidelines for correct and effective application of such methods are provided in order to gain detailed insight into the information governing the decision making process. We have obtained a number of textual reports from the banks regarding the information

DOI: 10.4018/978-1-4666-3886-0.ch025

collected from SMEs during the credit application/evaluation process. These are used as the basis for generating a synthetic XML database that partially reflects real-world scenarios. A tree mining method is applied to this data to demonstrate the potential of the proposed method for credit risk assessment.

INTRODUCTION

The emerging need for methods of credit risk assessment for Small Medium Enterprises' loan applications presents a unique challenge to the knowledge discovery and data mining field. The present credit scoring methods are considered not viable for SMEs since they are constructed from characteristics and risks pertaining to large scale business. In addition, SMEs are known for their imprecise management style, having non-systematic bookkeeping and organization of the business. This leads to a lack of valid and reliable financial information in traditional form (Berger, Klapper, & Udell, 2001; Berger & Udell, 1995) which is currently needed for the assessment of loan applications. In order to overcome the problem, loan staffs are required to collate data using a qualitative data collection method, namely interviews and observations. Therefore, a good portion of information on loan applications is available in a qualitative rather than quantitative form.

The abundant studies on credit scoring have contributed to credit risk methods being constructed using statistical and machine learning techniques. Aside from these mainstream techniques, our survey of the existing literature shows that a small number of researches have conducted studies using hybrid methods. Although each method shows respectable performance in classifying good and bad loan applications, each has inherent weaknesses. This, among others, is due to the fact that they are constructed using quantitative data which results in limited applicability of such a method in the real world of SMEs. Recent studies on credit risk assessment of SMEs highlight the necessity of incorporating qualitative information into the method (e.g., Dinh & Kleimeier, 2007). The level of qualitative data on SMEs loan applications is

variable in both quality and quantity. There are elements which could impact upon decisions regarding loan applications which are conspicuously higher in qualitative nature than others; these are goodwill, competency and integrity. These three characteristics require adequate elaborations since answers to these questions can only be understood by inference rather than a direct response.

This qualitative information is mainly available in free form text, which poses additional complications as most of the well developed and explored statistical and machine learning (data mining) methods are applied mainly to relational data with a well-defined structure. The task at hand is to develop a technique that incorporates and analyses qualitative information in tandem with quantitative information so that it accurately discloses applicants' credit risks. We propose a way to capture the qualitative information in a domain oriented way by defining an XML based template. We will show how the relevant information from the documents used by the banks for assessing credit risk for SME loan applications can be effectively captured using the proposed template. Within this context, preliminary results of the pre-defined XML template that are generated from a small number of textual document instances will be presented.

The main problem in association rule mining of semi-structured documents such as XML, is that of frequent pattern discovery, where a pattern in this case corresponds to a subtree. This is known as the *frequent subtree mining* problem, in which given a tree database T_{DB} and minimum support threshold (σ), the goal is to find all subtrees that occur at least σ times in T_{DB} . Driven by different application needs, several frequent subtree mining algorithms have been proposed in the literature that can mine different subtree types using different

36 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/incorporating-qualitative-information-credit-risk/75979

Related Content

Fiscal Policy and Social Optimization for Developing Nations: Some Thoughts in the Digital Era

Jose Manuel Saiz-Alvarez and Guillermo Calleja-Leal (2020). *Handbook of Research on Social and Organizational Dynamics in the Digital Era* (pp. 292-304).

www.irma-international.org/chapter/fiscal-policy-and-social-optimization-for-developing-nations/234546

Website Design

Stephen Burgess, Carmine Carmine Sellitto and Stan Karanasios (2009). *Effective Web Presence Solutions for Small Businesses: Strategies for Successful Implementation* (pp. 168-196).

www.irma-international.org/chapter/website-design/9245

Strategic Behavior in Family Firms: A Theoretical Insight Into Dynamic Familiness

Ismael Barros, Juan Hernangómez and Natalia Martín Cruz (2022). *Research Anthology on Strategies for Maintaining Successful Family Firms* (pp. 974-996).

www.irma-international.org/chapter/strategic-behavior-in-family-firms/288299

Social Skills and Competencies as the Driving Force of SME Development in Russia

Natalya Totskaya (2020). *Entrepreneurial Development and Innovation in Family Businesses and SMEs* (pp. 59-74).

www.irma-international.org/chapter/social-skills-and-competencies-as-the-driving-force-of-sme-development-in-russia/257087

Which Competitive Strategy Fits Better to Different Family-Business Profiles?: A Configurational Approach

Tomás F. González-Cruz and Norat Roig-Tierno (2020). *Competitiveness, Organizational Management, and Governance in Family Firms* (pp. 127-146).

www.irma-international.org/chapter/which-competitive-strategy-fits-better-to-different-family-business-profiles/241138