

An integration method combining Rough Set Theory with formal concept analysis for personal investment portfolios

Jhieh-Yu Shyng^{a,b,*}, How-Ming Shieh^a, Gwo-Hshiung Tzeng^{c,d}

^a Department of Business Administration, National Central University, No. 300, Chung-da Rd., Chung-Li City 320, Taiwan

^b Department of Information Management, Lan-Yang Institute of Technology, No. 79, Fu-Shin Rd, To-Chen, I-Lan 621, Taiwan

^c Department of Business and Entrepreneurial Management, Institute of Project Management, Kainan University, No. 1, Kainan Rd., Luchu, Taoyuan 338, Taiwan

^d Institute of Management of Technology, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:

Received 2 January 2010

Received in revised form 3 April 2010

Accepted 3 April 2010

Available online 13 April 2010

Keywords:

Rough Set Theory (RST)

Formal concept analysis (FCA)

Data mining

Choosing behaviours

Personal investment portfolios

ABSTRACT

The classical Rough Set Theory (RST) always generates too many rules, making it difficult for decision makers to choose a suitable rule. In this study, we use two processes (pre process and post process) to select suitable rules and to explore the relationship among attributes. In pre process, we propose a pruning process to select suitable rules by setting up a threshold on the support object of decision rules, to thereby solve the problem of too many rules. The post process used the formal concept analysis from these suitable rules to explore the attribute relationship and the most important factors affecting decision making for choosing behaviours of personal investment portfolios. In this study, we explored the main concepts (characteristics) for the *conservative portfolio*: the stable job, less than 4 working years, and the gender is male; the *moderate portfolio*: high school education, the monthly salary between NT\$30,001 (US\$1000) and NT\$80,000 (US\$2667), the gender is male; and the *aggressive portfolio*: the monthly salary between NT\$30,001 (US\$1000) and NT\$80,000 (US\$2667), less than 4 working years, and a stable job. The study result successfully explored the most important factors affecting the personal investment portfolios and the suitable rules that can help decision makers.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Real-world data may consist of incomplete and inconsistent information. We can process uncertain and/or incomplete information when the information is discovered knowledge. Data pre processing techniques can improve the quality of the data, accuracy, and the efficiency mining process. Since quality decisions must be based on quality data, data pre processing is an important step in the knowledge discovery process. Data mining [35] generates decision rules that can provide business managers with information about the competition in the market.

Research concerning attitudes towards personal wealth has increased in recent years. A well-designed financial plan can help customers achieve good asset allocation and meet their needs. However, few papers have been published on the topic of personal investment portfolios. The most important paper to submit the idea of the choice of portfolio was Markowitz, in 1952 [13]. The personal investment portfolio has been applied to many fields,

such as the behavior of financial services consumers [7], management of personal finances [17], retirement plans [6], and the assessment of the impact of customer satisfaction and relationship quality on customer retention [8]. In the paper of Keng and Hwa [10], they propose the residential property as an important component in a household's overall wealth.

The personal investment portfolio belongs to human knowledge which is a natural language. The natural language (or ordinary language) describes as general-purpose communications including speech, writing, or sign language for human in Wikipedia. Machine learning techniques are used to deal with uncertain data in natural language processing. The statistical natural language processing is mainly technology used for machine learning and data mining which both are fields of artificial intelligence.

The fuzzy set and the rough set theories are particularly adequate for the analysis of various data types, especially dealing with inexact, uncertain or vague knowledge. From a computational perspective, this study proposed the Rough Set Theory (RST), which is a rule-based decision-making technique that was developed by Pawlak [14]. Numerous applications of RST are presented in various scientific domains which have more details in the next section.

RST was used to analyze data contents and data features in this study. The results of RST are presented in the form of classification or decision rules derived from a set of data. It is also presented in

* Corresponding author at: Department of Business Administration, National Central University, No. 300, Chung-da Rd., Chung-Li City 320, Taiwan.

E-mail addresses: shyng@mail.ft.edu.tw, shyng@mgt.ncu.edu.tw (J.-Y. Shyng), hmsieh@mgt.ncu.edu.tw (H.-M. Shieh), ghtzeng@cc.nctu.edu.tw, ghtzeng@mail.knu.edu.tw (G.-H. Tzeng).

the form of “if... , then...” decision rules which seems to be more understandable for decision support.

The rule selection indices are the support objects of a rule, the compact of a rule, and the accuracy of a rule. It is a useless decision rule with only one support object due to decrease of the decision precision. Too many unqualified rules will decrease the decision precision. RST generates many rules and some of them have the same strength rate and the same number of support objects. These factors make it very difficult for decision makers to choose suitable rules. This study set up a pruning process which is the support object as a user-defined threshold. In this study we set up a decision rule with only one support object as threshold. This threshold can help to select the suitable rules in order to solve the problem of too many decision rules and to improve the decision precision.

Decision rules are the major information source for decision makers to do the data analysis. However, to explore the knowledge among rules is not an easy way. In this study we used formal concept analysis (FCA) to aggregate the suitable decision rules to provide the prior information for decision makers. The lattice diagram was provided by the FCA in order to gather the decision rules, to construct the concept and to explore the relationship among attributes. The FCA provides the mathematical theory, which belongs to algebra and is a branch of lattice theory.

The FCA is a theory of data analysis that constructs the conceptual structures among data sets. It was introduced by Wille [4] and has since grown rapidly. The FCA is a duality notion that can often be observed between two types of items that relate to each other in an application, such as objects, and attributes, or documents and terms. Conceptual relationships are discussed by the data matrices (contexts) formed by attributes and objects. Another, a mathematical model allows us to study mathematically the representation of conceptual knowledge.

In RST, the data for analysis are described by information system (U, A, R) , which corresponds to the formal context in FCA and consists of universe U , attributes set A , and the relation R between U and A . RST and FCA are two complementary mathematical tools for data analysis. Knowledge processing and data analysis always uses concepts to elaborate interpretations of given data and information.

In this study, we use two steps to perform the data analysis. The first step is pre process, which focuses on the problem of many decision rules, and sets up a rule threshold using the support object of a rule to reduce the number of rules. The main purpose is to find the suitable rules. The second step is post process, which creates additional values on those suitable rules by the FCA in order to find the relationship among attributes and to construct the conceptual structures among data sets. One of the greatest benefits is that the decision maker can have a first insight before data analysis. The complete process steps are shown in Fig. 1.

For this study, a questionnaire was designed to investigate personal investment portfolios, using real cases of investors in Taiwan as the basis of the empirical study. The questionnaire considered the factors affecting decision making, such as gender, age, the number of family members, monthly income [7,17], and participants' basic data (such as Marriage Status, Education, Number of Working Years, Professional Status), which may serve as a basis for understanding their needs.

The results of the study identify three types of personal investment portfolios: a conservative portfolio, a moderate portfolio, and an aggressive portfolio. The main general concepts (characteristics) of investors who choose conservative portfolios are having a stable job, low working years, and male; investors who have aggressive portfolios have a higher income and more working years; investors who have moderate portfolios are usually high school-educated and male. More details are presented later.

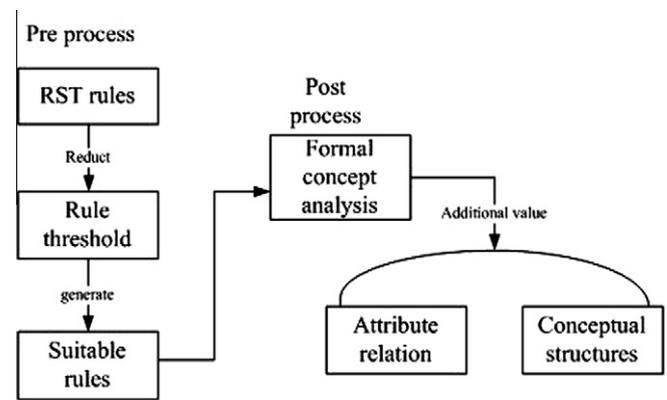


Fig. 1. Study process steps.

In this study, the most important factors affecting the personal investment portfolios were the job type (stable or non-stable), the monthly salary, and education, which carried the greatest affects on the conservative portfolio, moderate portfolio and aggressive portfolio, respectively.

The remainder of this paper is organized as follows. Section 2 describes the concepts to be used in this study. In Section 3, a real case of personal investment portfolio is presented to show the process of this study. In Section 4, we present our conclusions.

2. Concepts about this study

In this section, we briefly introduce RST and FCA, which are used in analyzing the personal investment portfolio. In Section 2.1, the RST is described. In Section 2.2, the FCA is presented.

2.1. Rough set theory and background

RST is a tool for processing uncertain and incomplete information; in this theory, the lower and upper approximations of an arbitrary subset of universe U are the basic operators. In 1982, Pawlak designed RST as a tool to describe the dependencies between attributes, evaluate the significance of the attributes, and deal with inconsistent data. Both fuzzy set theory [28] and RST are used with the indiscernibility relation and perceptible knowledge. The major difference between them is that RST does not need a membership function. A detailed discussion of RST can be found in Walczak and Massart [23]. RST has been applied to the management of a number of issues, including medical diagnosis, engineering reliability, expert systems, empirical study of insurance data [20], machine diagnosis [29], business failure prediction [1], activity-based travel modelling [25], and data mining [2,9,21,34]. Another paper discusses the preference-order of attribute criteria needed to extend the original RST, such as sorting, choice, and ranking problems [5].

RST applies the indiscernibility relation and data pattern comparison based on the concept of an information system with indiscernible data, where the data is uncertain or inconsistent. The data is grouped into classes called elementary sets. More detailed information regarding attributes can be found in the works of Swiniarski and Skowron [22] and Polkowski [18]. The objects in a class may have a relationship with the corresponding features/attributes, and expert knowledge is used to process attribute extraction. Each elementary set is independent of the others. We can extract knowledge from each elementary set used in the real world.

In this section, we will discuss topics, such as the indiscernibility relation, classification, set approximation, reduct and core attribute sets, and decision rules related to RST.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات