European Journal of Operational Research 201 (2010) 601–607

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Using FSBT technique with Rough Set Theory for personal investment portfolio analysis

Jhieh-Yu Shyng [a,d,*], How-Ming Shieh [a], Gwo-Hshiung Tzeng [b,c], Shu-Huei Hsieh [d]

[a] Department of Business Administration, National Central University, No. 300, Chung-da Road, Chung-Li City 320, Taiwan
[b] Department of Business and Entrepreneurial Administration, Kainan University, No. 1, Kainan Road, Luchu, Taoyuan 338, Taiwan
[c] Institute of Management of Technology, National Chiao Tung University, No. 1001, Ta-Hsueh Road, Hsinchu 300, Taiwan
[d] Department of Information Management, Lan-Yang Institute of Technology, No. 79, Fu-Shin Road, To-Chen, I-Lan 621, Taiwan

A R T I C L E I N F O

Article history:
Received 4 September 2007
Accepted 20 March 2009
Available online 29 March 2009

Keywords:
Asset allocation
Asset management
Forward Search and Backward Trace (FSBT)
Investment portfolio
Rough Set Theory (RST)

A B S T R A C T

This study proposes a novel Forward Search and Backward Trace (FSBT) technique based on Rough Set Theory to improve data analysis and extend the scope of observations made from sample data to solve personal investment portfolio problems. Rough Set Theory mathematically classifies data into class sets. The class set with the most objects may generate one decision rule. The rules generated from RST are rough and fragmented, that are very difficult to interpret the information. An empirical case is used to generate more than 85 rules by the RST method in comparison with FSBT method which only generated 14 rules. This result can show our proposed method is better than traditional RST method based on class sets that contain the most objects. Much of human knowledge is described in natural language. It is a very important thing to convert information from computer databases into normal human language. Sample data taken from features with the same backgrounds are used to compile different portfolios that investment companies and investment advisors can employ to satisfy the investor' needs. The method not only can provide decision-making rules, but also can offer alternative strategies for better data analysis. We believe that the FSBT technique can be fully applied in research on investment marketing.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

A well-designed financial plan can help to achieve good asset allocation and meet customer needs. Indeed, asset management is closely tied to personal experience and behavior. For example, in an inflationary market, investment capital decreases as personal expenses increase.

In recent years, research concerning attitudes towards personal wealth has increased. Moreover, there has been an increase in the public's interest in wealth creation, which is perhaps best reflected in the burgeoning literature on financial matters. With regard to financial hardship, research suggests that individuals' past experiences usually affect their attitudes towards making investments. However, there is now a trend toward relying on financial services as an effective tool for increasing personal wealth.

Assessing financial services is a complex task because the consumer has to evaluate the features of product before it has actually been consumed. For example, when consumers adopt a particular investment plan, they cannot really evaluate the quality of the outcome until the investments mature. Hence, they may be forced to make a decision before assessing the result.

Among financial institutions, there is a growing preoccupation with customer retention and relationship marketing – essentially understanding the behavior of consumers after the initial purchase has been made, and focusing on how to foster a profitable relationship with the clients. This is crucial to the success of financial institutions, and the approach recognizes the ongoing nature of the relationship and the longevity of many financial products.

Prior to making purchase decisions, consumers can gather information from a variety of sources. The factors that affect decisions associated with personal asset allocation are the risk level and return revenue of investment products, the timing strategy (the time of buying and selling products), and the portfolio. Additionally, individuals may have different investment needs, reflecting their personal backgrounds and life experiences as well as their individual personalities. These factors can make personal asset allocation decisions very difficult.

To date, there have been relatively few attempts to develop models that explain consumer decision processes specifically in the context of financial services. A considerable amount of theoretical and empirical work exists relating to how businesses make

* Corresponding author. Address: Department of Business Administration, National Central University, No. 300, Chung-da Road, Chung-Li City 320, Taiwan.
E-mail addresses: shyng@mgt.ncu.edu.tw, shyng@mail.fit.edu.tw (J.-Y. Shyng), hmshieh@mgt.ncu.edu.tw (H.-M. Shieh), ghtzeng@cc.nctu.edu.tw, ghtzeng@mail.knu.edu.tw (G.-H. Tzeng), hsieh@mail.fit.edu.tw (S.-H. Hsieh).

0377-2217/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2009.03.031

financial decisions for clients generally, rather than how they personalize their services for a single customer. From the buyer's perspective, the decision-making process can be roughly divided into four components: problem recognition, information search, evaluation of alternatives, and purchase decision. The decision-making process may also be affected by one's past life experience and personality. The process of selecting an appropriate investment portfolio can be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performance of available securities. The second stage starts with relevant beliefs about the future performance of various investment products and ends with the choice of portfolio (Markowitz, 1952). Many papers have been published on this topic; specific subjects include the behavior of financial services consumers (Harrison, 2003), management of personal finances (Teichman et al., 2005), retirement plans (Hanisch, 1994), personalization of intelligent financial decision support systems (Palma-dos-Reis and Zahedi, 1999), development of an intelligent system for personal and family financial services (Chieffe and Rakes, 1999), and assessment of the impact of customer satisfaction and relationship quality on customer retention (Hennig-Thurau and Klee, 1997). A number of studies focus on quantification of the problem by streamlining all the parameters and applying statistical tools to analyze the data. Rough Set Theory (RST) is a method for discovering knowledge from ambiguous information. Developed by Pawlak (1982, 1984), RST is a rule-based decision-making technique that can handle crisp data sets and fuzzy data sets, without the need for a pre-assumption membership function, as required by fuzzy theories (Zadeh, 1965). Membership is not the main concept in the Rough Set Theory. Rough sets can deal with indiscernibility knowledge and fuzzy sets deal with vagueness knowledge. The results of the Rough Set Theory are composed of classification and decision rules which derived from a set of examples. More comparison details between fuzzy set and rough set theories can refer the paper of Walczak and Massart (1999). RST has become a hot topic because of its application to knowledge discovery in real-world databases or warehouses.

However, RST generates many rough rules that are not easy to apply. In this empirical study, there are more than 85 rules generated by the RST method. Thus, we propose a new search method called Forward Search and Backward Trace (FSBT), which extends the search process after the RST technique finds the decision class set with the most objects and only generate 14 rules in FSBT method. That set represents the desired answer or model, and we use its components as criteria in a forward search to find matching objects (called Target objects) that have the same components as the decision class set with the most objects. We can then discover a consumer's consumption behavior from those components (the customer's backgrounds). It is a very important thing to convert information from computer databases into normal human language and much of human knowledge is described in natural language. Complete information for a rule is an important thing for Decision Maker to do the decisions.

The rules generated from RST are rough and fragmented, that are very difficult to interpret the information. However, the rough set approach can be considered as a formal framework for discovering facts from imperfect data. The advantage of FSBT is that, compared to classical RST, it can cull more information from the same backgrounds/personalities data sample. Classical RST only generates rough rules that have the best lower approximation rate to perform an analysis. This result for data utilization is very inefficient. However, the FSBT method can expand the observation scope and reduce data utilization inefficiency. Financial advisors can derive more investment combinations with FSBT applications. In reality, investment experts can review a customer with the personal characters/backgrounds, estimate which investment portfolio is suitable for the customer and give adequate suggestions by the FSBT method. Moreover, the investment experts can perform special analysis and provide further service to the special customers (the outlier objects) supplied by the FSBT method.

In this paper, we focus on identifying different types of information derived from the same characters/backgrounds to create more personalized investment portfolios for satisfying the investors' needs and helping them increase their personal wealth. We use RST for feature classification, and then try to determine the priority of investments at different levels. The results may help investment companies or investment advisors propose suitable investment products to fulfill their customers' needs. There have been relatively few studies on the use of RST for personal investment analysis. For this study, we designed a questionnaire to investigate personal investment portfolios. Using real cases of investors in Taiwan as the basis of the empirical study, we tried to determine the participants' priorities when they make personal investment portfolio decisions. The questionnaire considered the factors that affect decision-making, such as sex, age, and the number of family members; monthly income (Harrison, 2003; Plath and Stevenson, 2005); the nature of the investment products; and participants' basic data, which may serve as the basis for understanding their needs.

The results of the study demonstrate the efficacy of the proposed FSBT method, and identify three types of personal investment portfolio: conservative portfolio, moderate portfolio, and aggressive portfolio. Investors who choose conservative portfolios are generally single college graduates under the age of 29, with less than four years work experience and a monthly salary below US$900. The majority of people with an aggressive portfolio are married, female college graduates under the age of 39, with 15–19 years work experience and a monthly salary below US$2424. Those with a moderate portfolio are generally single, under 29 years of age, with a monthly salary between US$900 and US$2424. The investment priorities for conservative portfolio consumers are bank deposits, insurance policies, and houses. The priorities of those with moderate portfolios are bank deposits, houses, and insurance, while those with aggressive portfolios prefer houses, bank deposits, and stocks.

The remainder of this paper is organized as follows. In Section 2, we discuss personal investment portfolios. In Section 3, the methodology of FSBT combined with RST is described. In Section 4, we consider a real case of a personal investment portfolio to explain the process of the proposed method. Then, in Section 5, we present some concluding remarks. To distinguish between RST and FSBT, classical RST is denoted as RST throughout the paper.

2. Personal investment portfolio

In an investment portfolio, the factor with the most impact on investment revenue is asset allocation. Hence, investors should modify their investment portfolios depending on their needs at regular intervals. An investment portfolio is composed of many investment products that must be managed effectively in order to increase personal wealth.

Investment products can be divided into two categories: risk products and non-risk products. Risk investments include stocks, mutual funds, foreign exchange, land, houses, and investment insurance. Non-risk investments include bank deposits, traditional insurance (such as life insurance, AD&D), and government bonds. To simplify the analysis and improve the precision of decision-making, we only consider eight investment products: bank deposits, insurance products (traditional insurance and investment insurance), houses, land, stocks, mutual funds, and foreign exchange. Among the above-mentioned investment products, deposits and insurance are considered non-risk products; the other six belong to the category of risk products (see Supplementary material).

The proposed asset allocation model can be divided into five categories or types of personal asset allocation portfolio: conservative portfolio, moderately conservative portfolio, moderate portfolio, moderately aggressive portfolio, and aggressive portfolio. To simplify the analysis, we combine the portfolios into three types: non-aggressive (conservative) portfolio, moderate portfolio, and aggressive portfolio.

By summarizing media reports, percentages of personal investment targets can be proposed when compiling a portfolio. If the risk of investments reaches 80%, the portfolio is considered aggressive. In contrast, if a portfolio is composed of 100% non-risk investment products, it is deemed non-aggressive or conservative. Asset allocation portfolios that fall between the non-aggressive and the aggressive portfolios are considered moderate. These concepts of classification for personal investment portfolios can be shown as Figure 1 (see Supplementary material).

3. RST and FSBT methods

We now introduce RST and FSBT methods and explain the procedures used to analyze personal investment portfolios. Section 3.1 describes the history of RST, Section 3.2 details the RST algorithms, and Section 3.3 discusses the FSBT method. Decision rules can be approached by enhancing the approximations and accuracy of the decision rules of the classification and checking the quality of the classification, all of which can be used to extract suitable rules and describe the information system. In fact, the application of RST under real conditions usually produces too many rough rules; therefore, it is difficult to select a single optimal rule.

3.1. The history of RST

To use knowledge from a real-world database, we need to consider the stored records as well as the number of attributes that may be relevant to the decision that we need to make. For these reasons, RST has become a popular method for discovering knowledge and data mining in practical applications. The major advantage of RST is that it can be applied to inexact, uncertain, and ambiguous datasets.

Like RST, fuzzy set theory (FST) is used with the indiscernibility relation and perceptible knowledge. The major difference between FST and RST is that the latter does not make any presumptions or require a priori knowledge about the data, such as the grade of the membership function in FST (Zadeh, 1965).

Developed by Pawlak (1982, 1984, 2004), RST has evolved into a rule-based decision-making technique that is essentially a mathematical tool for dealing with ambiguous or uncertain information. It has been applied in many fields, such as medical diagnosis, engineering reliability, expert systems, machine diagnosis (Zhai et al., 2002), business failure prediction (Beynon and Peel, 2001; Dimitras et al., 1999), activity-based travel modeling (Witlox and Tindemans, 2004), decision logics languages for rule representation (Fan et al., 2007), the solution of linear programs (Azibi and Vanderpooten, 2002), data mining (Hu et al., 2003; Chan, 1998), α-RST (Quafafou, 2000), feature selection and recognition (Swiniarski and Skowron, 2003), as well as sorting, choosing, and ranking problems (Greco et al., 2001). And Sawicki and Zak (2009) study is a process of technical diagnostic applied to a fleet of vehicles which Classical rough set (CRS) theory is compared with the dominance-based rough set (DRS) approach. RST is useful for exploring data patterns and determining the relative importance of each attribute with respect to its output.

3.2. The RST algorithms

Pawlak designed RST as a tool to describe the dependencies between attributes, evaluate the significance of the attributes, and deal with inconsistent data. We now discuss topics related to RST, such as the indiscernibility relation and classification, approximation accuracy; reduct and core attribute sets, and decision rules.

Definition 1 (*Information systems*). Given a questionnaire model QM (an information system), $QM = (U, A, V, \rho)$; $U = \{x_1, x_2, \ldots, x_n\}$, where U denotes the universal object sets of QM; A = {features/attributes}, for example, $A = \{a_1, a_2, a_3\}$, A represents the model's attribute sets, consisting of attributes $\{a_1, a_2, a_3\}$; and $V = \cup_{a \in A} V_a$, is a set of values of the attributes.

Let $\rho{:}U \times A \to V$ be a description function; and let ρx be the description of x in QM, where $\rho(x,a) \in V_a$ for each $a \in A$ and $x \in U$ (Pawlak, 2002).

Walczak and Massart (1999) expressed that a knowledge representation system containing the set of condition attributes (denoted as CA) and the set of decision attributes (denoted as DA) is called a decision table.

Definition 2 (*Indiscernibility relation and classification*). Pawlak expressed "any subset $X \subseteq U$ of the universe will be called a concept or a category in U" that is concepts to form a partition (classification) of a certain of universe U; assume a family $Y = \{X_1, X_2, \ldots, X_m\}$ is a family of nonempty sets (classification) that $X_i \subseteq U$, $X_i \neq \emptyset$, $X_i \cap X_j = \emptyset$ for $i \neq j$, $i,j = 1, 2, \ldots, m$ and $\cup_{i=1}^{m} X_i = U$.

Any subset B of A determines a binary relation IND(B) on U, which we call an indiscernibility relation, and define it as $a \in B$, if $\rho_{x_1}(a) = \rho_{x_2}(a)$ for every $a \in A$. The equivalence class of IND(B) is called an elementary set (of atoms) of QM. The partition of U with respect to IND(B) is denoted by U/IND(B). Objects grouped in the same class are called elementary sets, and the process is called classification. Thus, any x_i of U can be induced so that the value sets of attributes represented in B are in the same class. A set represents the smallest discernible group of objects, and the construction of elementary sets is an important step in the classification with rough sets. The classification that processes CA and DA, generates condition and decision classes.

Definition 3 (*Attribute dependence and approximation accuracy*). Let $QM = (U, A, V, \rho)$ be an information system, and $A = \{a_1, a_2\}$ represents the model's attribute sets. The attribute dependence can be defined as: (1) ($a_1 \to a_2$) iff $a_1 \subset a_2$, attribute a_2 is said to be dependent on attribute a_1 in QM; and (2) iff neither $a_1 \to a_2$ nor $a_2 \to a_1$, the attributes a_1, a_2 are said to be independent of QM in Pawlak (1984). Thus, we can induce IND(A) = IND($A - a_2$), and a_2 is a superfluous attribute. We can remove the superfluous attributes to simplify the information set and generates diagnostic values. The way to check the dependency of a set of attributes and find the superfluous attributes that can be removed; attributes are checked sequentially while computing the number of each elementary set. The attribute is defined as a superfluous when the number of the elementary set is the same as the original set and the remaining attributes are considered indispensable.

Let any subset $X \subseteq U$, and R be an equivalence relation and x_i expresses objects $x_1, x_2, \ldots, x_n$, where $i = 1, 2, \ldots, n$, then $\underline{R}X = \{x \in U{:}[x]R \subseteq X\}$, is the lower approximation (also called the R-lower approximation of X); $\overline{R}X = \{x \in U : [x]R \cap X \neq \emptyset\}$ is the upper approximation (also called the R-upper approximation of X); $Bnd_R(X) = \overline{R}X - \underline{R}X$, the boundary region of X (also called the R-boundary of X) that the objects are inconsistent or ambiguous. If a family $Y = \{X_1, X_2, \ldots, X_m\}$ of nonempty sets (classification), then $\underline{R}Y = \{\underline{R}X_1, \underline{R}X_2, \ldots, \underline{R}X_m\}$ and $\overline{R}Y = \{\overline{R}X_1, \overline{R}X_2, \ldots, \overline{R}X_m\}$, are called the R-lower and R-upper approximation of the family Y, respectively. Two measures about the inexactness of approximate classifications described as below. The first one is the accuracy of approximation of Y by R is derived from the computation of the intersection rate between the lower and upper approximations, and is mathematically defined as follow:

$$\alpha_R(Y) = \frac{\sum card(\underline{R}X_i)}{\sum card(\overline{R}X_i)}, \text{where } card \text{ means cardinality of a set.}$$

The percentage of possible correct decisions described as the accuracy of classification when classifying objects employing the knowledge R.

The second is the quality of approximation of Y by R is mathematically as follow:

$$\tau_R(Y) = \frac{\sum card(\underline{R}X_i)}{card(U)}.$$

The percentage of objects described as the quality of the classification which can be correctly classified to classes of Y employing knowledge R.

Definition 4 (*Reduct and core attribute sets*). Two fundamental concepts of Rough Set Theory are reduct and core attribute sets. The most precise way of discerning object classes are reducts which are the minimal subsets provided that the object classification is the same as with the full set of attributes and the core is common to all reducts.

The goal of reducts is to improve the precision of decisions that the reducts process for attributes reduces elementary set numbers. The complete set of attributes is called a reduct attribute set. There may be more than one reduct attribute set in an information system, but intersecting a number of reduct attribute sets yields a core attribute set. The reduct attribute set affects the process of decision-making, and the core attribute is the most important attribute in decision-making.

In an information system, some attributes may be redundant and useless. If removed them without affecting the classification power of attributes (Pawlak, 1984). Given an attribute space $A = (CA,DA)$. Let $RED(B) \subseteq A$; $RED(B)$ be the reduct set composed of a set of attributes B, which affects the process of decision-making that is a minimal set of attributes. There may have several reduct attribute sets. The intersection of all reduct attribute sets is the core attribute set, which is the most important attribute in the decision-making process, $COR(C) = \cap RED(B)$ which $COR(C)$ is the core composed of a set of attributes C. Applying the reduct set to the model, we can induce the decision rules.

Definition 5 (*Decision rules*). Given an attribute space $A = (CA,DA)$, where $CA \neq \emptyset$, $DA \neq \emptyset$; then $DA \cap CA = \emptyset$, and $DA \cup CA = A$, which are the elements of the decision table. Assumes an indiscernibility relation, IND(DA); if objects that have the same IND(DA) are grouped together and called decision elementary sets (decision classes). The reduct condition attribute sets maintain the important relationships with decision classes. Due to the functional dependencies between conditions and decision attributes, a decision table may also be seen as a set of decision rules. The syntax can use the "if..., then..." rule to specify as "*if..., then...*". The syntax of the rule is as follows:

If $f(x,a_1)$ *and* $f(x,a_2)$ *and ... and* $f(x,a_k)$, *then* x belongs to ds_1 or ds_2 or ds_n, where $\{a_1,a_2,\ldots,a_k\} \subseteq CA$ are condition attributes and $\{ds_1,ds_2,\ldots,ds_n\} \subseteq DA$ are decision classes. If $n = 1$, then the rule is exact; otherwise, it is approximate or ambiguous (Greco et al., 2001).

According to Pawlak (2002), a decision rule in QM is expressed as $\Phi \to \Psi$, where Φ and Ψ are conditions and decisions of the decision rule, respectively (read as: if Φ then Ψ); $\sigma_{QM}(\Phi,\Psi) = \text{supp}_{QM}(\Phi,\Psi)/card(U)$ is the strength of the decision rule $\Phi \to \Psi$ in QM, where the number $\text{supp}_{QM}(\Phi,\Psi) = card(\|\Phi \wedge \Psi\|_{QM})$ will be called the *support* of the rule $\Phi \to \Psi$ in QM; and $card(U)$ is the cardinality of set U. This implies that a stronger rule will cover more objects and the strength of each decision rule can be calculated in order to decide the appropriate rules. An induction algorithm to generate: (1) a minimal rule set that can cover all objects, (2) a rule that can cover all possible rules, and (3) the strongest rule that can cover many objects was proposed by Dimitras et al. (1999).

RST usually assumes that there is only one decision attribute. However, in this study we focus on the problem of the multi-attribute decision. Let $QM = (U,A,V,\rho)$, U be the universal objects of QM, and A the model of attribute sets that can be divided into two parts (CA and DA), for example, where $CA = \{a_1,a_2,a_3\}$ and $DA = \{d_1,d_2,\ldots,d_8\}$ are condition attribute set and decision attribute set, respectively. The type of decision table expressed in the terms of attributes CA can be expressed in the terms of attribute DA.

It is difficult to group data into classes if most data patterns are unique and because the attributes of combination values form a separate class (Shyng et al., 2007). Each unique data class forms an individual class, which reduces the approximation accuracy and makes it more difficult to interpret the decision rule. The same applies to multi-attribute decisions (Shyng et al., 2006). However, RST generates many rough rules that are not easy to apply. Therefore, in this study, we propose a new search method called Forward Search and Backward Trace (FSBT), which extends the search process after the RST technique finds the decision class set with the most objects. That set represents the desired answer or model, and we use its components as criteria in a forward search to find matching objects (called Target objects) that have the same components as the decision class set with the most objects. We can then discover a consumer's consumption behavior from those components (the customer's backgrounds). In the next section, we present the Forward Search and Backward Trace method.

3.3. The FSBT method

The FSBT method extends the original RST method to discover the characteristics of objects (items) from information system. Given an attribute space $A = (CA,DA)$, where CA (condition attributes) are used to describe the characteristics of objects, and DA (decision attributes) define an elementary set according to the condition attributes. Applied to a real case, classical RST yields very rough results. Therefore, we apply the FSBT method to improve the precision of the analysis and enlarge the observation scope. The basic idea of FSBT can explored from chart in Figure 2 (see Supplementary material). The FSBT method employs the following six-step algorithm.

Step 1: Data set
INPUT

$QM = (U,A,V,\rho)$; $A = \{\text{features/attributes}\} = \{CA, DA\}$; $CA = \{c_i;\ i = 1,2,\ldots,nn\}$;
$DA = \{d_j;\ \ j = 1,2,\ldots,mm\}$; $V(CA) = \cup_{c_i \in A} V_{c_i}$; $i = 1,2,\ldots,nn$; $V(DA) = \cup_{d_j \in A} V_{d_j}$; $j = 1,2,\ldots,mm$;
$DS = \{ds_w;\ \ w \in W;\ \ w = 1,2,\ldots,m\}$; $CS = \{cs_t;\ \ t \in T;\ t = 1,2,\ldots,n\}$;
CA represents the condition attributes sets; thus, $CA \subset A$; DA represents the decision attributes sets; thus, $DA \subset A$; $V(CA)$ is a set of values of the condition attributes sets; $V(DA)$ is a set of values of the decision attributes sets; DS represents the

decision class sets; *CS* represents the corresponding condition class sets in the decision class sets; *X* is a subset of *U*.
OUTPUT
Portfolio types: Aggressive, Moderate, and Conservative.
Step 2: Find decision class set
Find *DS*
From the decision class sets (*DS*), select three sets, ds_1, ds_2, ds_3, to represent the conservative, moderate, and aggressive types of portfolio, respectively. Note that ds_1,ds_2, ds_3, are the class sets in *DS* that have the most objects.

Step 3: Forward search
Find the Selected objects, i.e., the corresponding objects in ds_1, ds_2, ds_3. The set with the corresponding objects is denoted as *CS*. The corresponding condition class sets, cs_1, cs_2, cs_3, represent ds_1, ds_2, ds_3, respectively.

Step 4: Extend approach
For each (ds_i; $i = 1, 2, 3$)
Begin
    $criteria \leftarrow Vx_i(CA)$; $\forall x_i \in ds_i$
While ! END{$criteria \leftarrow Vx_i(CA)$}
Begin
    While ! END(x_j; $j = 1,2,\ldots,n, x_j \in U$)

Step 5: Backward trace
    Begin
        if($Vx_j(CA) == Vx_i(CA)$) then //Find the Target objects
            Output $Vx_j(DA)$ //Output the decision data
            End
        End
    End

Step 6: Data analysis
Compared those output data.

To sum up, *Step 1*, the data set (input/output) step sets up the terms for the algorithm. *Step 2* finds the decision class set *DS*, from which we choose three decision class sets, ds_1, ds_2, ds_3; that is, the sets with the most objects used to represent the three portfolio types. *Step 3*, *Forward Search*, finds the corresponding objects in the ds_1, ds_2, ds_3. We call these objects *Selected objects*, each of which acts as a criterion for the search in *Step 4*, i.e., the *Extend Approach* step. In *Step 5*, the *Backward Trace* searches the objects matched based on the criteria in *Step 4* and outputs those that match the object. The objects matched in *Step 5* are called *Target objects*. Finally, *Step 6*, the data analysis step uses the experts' knowledge to explore the knowledge for the characteristics of the portfolio from the final result.

In fact, the main concept of FSBT is related to the lower approximation and upper approximation. The decision class set with the most objects is found in *Step 2* by choosing the decision class set with the best lower approximation rate. Based on the chosen lower approximation, the class set is reversed to find the corresponding upper approximation class set. Then, in the final step, based on the observed information, the corresponding objects of the upper approximation class set are selected, as shown in Figure 3 (see Supplementary material). In the next section we present our empirical study to demonstrate the proposed FSBT method, which is based on the FSBT method. The more detail concepts about the FSBT method shown in Table C.1 of Appendix C (see Supplementary material).

4. Empirical study of the FSBT method compared with the classical RST method for personal investment portfolio selection

Under highly competitive conditions, the best way to access markets and enlarge market share is to acquire needed information from potential customers through well-designed surveys. A successful business not only fulfills customers' needs, but also designs business strategies and/or measures to improve the firm's performance.

The questionnaire for personal investment portfolio analysis is discussed in the following.

4.1. Problem descriptions

According to the report of Taiwan's Directorate-General of Budget, Accounting and Statistics (Executive Yuan, 2005), the average family income and expenditure per household declined in the five years prior to the report. Average disposable income is increasing at a lower rate than the expenditure rate, and the savings rate is decreasing year by year. Many developed countries in Asia, such as Japan, are facing the same situation (Ohmae, 2006). The world economy has gradually evolved into two major groups: a small number of extremely rich countries and a large number of extremely poor countries, with the number of middle economic class countries declining sharply. The discrepancy between rich and poor is larger than ever. Therefore, determining how to increase wealth for a person with limited capital is an important issue.

In this study, we asked a series of questions to obtain information about consumers' investments in 2006, e.g., monthly salary, investment products as a percentage of total assets, and total investments. The participants' personal data, namely their gender, age, marital status, number of children, professional status, number of years in the workforce, and educational level was used to classify purchase intentions, based on the definition by Executive Yuan (2005).

4.2. Empirical processes

The questionnaires were distributed to investors in the North and Northeast districts of Taiwan. Data was collected based on a nominal scale. There were 200 valid questionnaires from a total of 221 received. The percentage of valid questionnaires is 90%. Among the valid respondents, there were 108 females and 92 males (see Supplementary material).

4.3. Empirical results and implications

4.3.1. Results

This study generated many rules for selecting personal investment portfolios by the classical RST approach. Several rough rules were extracted, as shown in Table 6 (see Supplementary material).

People with conservative personal investment portfolios tend to be single college graduates under 29 years old, with less than four years work experience and a monthly salary under NT$30,000 (US$900). The profile of a typical aggressive investment portfolio holder is as follows: married, female, college graduate, less than 39 years old, with 15–19 years work experience and a monthly salary under NT$80,000 (US$2424). The moderate personal investment portfolio is typically chosen by single college graduates less than 29 years old, with less than nine years work experience and a monthly salary between US$900 and US$2424. The results of our proposed method based on the decision class sets with the most objects are shown in Table B.2 of Appendix B (see Supplementary material).

4.3.2. Implications

For a person with a stable income who adopts a suitable retirement plan, the first investment priority is bank savings followed by housing in Taiwan. Housing is a basic necessity and represents the single larger investment next the bank saving in a personal investment portfolio. A conservative portfolio provides a stable rate of return with minimal risk. Furthermore, the global economic downturn, investors reduce the risk products investment and increase the conservative products investment corresponding with the investment experts' suggestions.

The average percentage of the investment target percentage for each portfolio of this empirical study results is shown in Table 7 (see Supplementary material). The highest investment target percentage is deposits, which is 65% in a conservative portfolio. The financial plans of conservative portfolio holders typically include deposits and insurance as well as small investments in higher risk products. The investment target percentage for the moderate portfolio is very even. The sum of the percentages of deposits, insurance, and houses is 62%, reflecting the mix of investment targets in a moderate investment plan with the items equally distributed. Investors with moderate portfolios are willing to take slightly more risk than investors with conservative portfolios. The highest investment target percentage is the aggressive portfolio, with only 26% invested in deposits and insurance and 45% invested in house.

4.4. Discussion

The FSBT approach selects the decision class set that contains the most objects and adopts it as a model for personal investment portfolios. The decision class set can generate one of the decision rules of classical RST, which means that a stronger rule covers more objects. In fact, a single object decision class set increases the number of decision rules, but degrades the decision precision (Shyng et al., 2007). Computation of the approximation accuracy is very important to decision rules extraction in classical RST, a process which provides the decision rules used by decision-makers (DM). Table B.1 (see Supplementary material) lists the decision class set that contains the most objects. The results listed in Table B.2 were generated using the FSBT approach based on the decision class set containing the most objects (Table B.1). The proposed FSBT method can extend the observation scope, from the 19 data samples generated via classical RST in Table B.1 to the 52 data samples generated via FSBT in Table B.2. Importantly, Table B.2 also reveals that the first priority in investment target selection is deposits followed by houses and insurance and up to 62% of data samples (investors) have conservative portfolio.

The results of FSBT reveal that young, single investors, with low income invest primarily in deposits, houses, and insurance, and always adopt conservative portfolios. According to this study, some young individuals invested more in property than deposits, given the expense of renting housing in modern cities. Paying off a mortgage is better than paying rent if the monthly mortgage equals the rent (or if it slightly exceeds the rent). However, to invest in property young people naturally need extra money to invest in the first place. Some investment advisors suggest that young people rent rather than buying a house, since this allows them to free up capital to make other investments that are likely to increase their wealth more rapidly than investing in residential real estate.

Different portfolios can be generated from the same background sample data, as shown in Table B.2. Thus, it can be assumed that portfolio differences result from environmental factors, including parental education, familial economic status, investor education level, and the influence of relatives and friends. These factors impact investment preferences.

The results listed in Table 7 demonstrate that most investors view deposits as their first choice investment vehicle, followed by houses and insurance. Clearly, most people are conservative in asset managing. However, based on the current situation, saving money in a bank is a poor investment because of low interest rates approaching zero in Taiwan. Generally, people spend primarily on living expenses (including monthly rent or mortgages) and keep the remainder in the bank for emergency spending.

Normally stocks and funds are the main components of an aggressive portfolio. Because of cultural differences, housing is the main investment target for holders of aggressive portfolios. The results of this study confirm the findings of Keng and Hwa (2004) and Norland (1988), demonstrated that housing is an important component of overall household wealth. For cultural reasons, Chinese people see house ownership as an essential component of wealth.

The advantage of FSBT is that it can cull more information than classical RST from the same data sample. Classical RST is very inefficient and FSBT method can expand the observation scope and reduce the inefficiency of data utilization. Classical RST only generates rough rules that have the best lower approximation rate to perform an analysis. The decision part of the rule thus defined the assignment of objects to specific classes. If the decision part of the rule indicates a single class then the rule is considered an exact decision rule, while otherwise it is considered an approximate rule (Sawicki and Zak, 2009). Financial advisors can derive investment combinations with FSBT applications. The proposed method enlarges the data scope by extending the original lower approximation set into the upper approximation set.

This empirical study generated more than 85 rules by the RST method in comparison to 14 rules with the FSBT method. This empirical study contains numerous class sets of one object (the ex-aequo class). For the reason, this empirical study attempts to discover the real facts of personal investment. In this study, there are eight investment products (decision attributes) for personal investment portfolio and the investment percentage (decision attribute value) for each investment product is a real number. The number of decision rules will increase as if too many decision attributes and too much unique data occurs at decision part. If too many decision rules are produced that do not benefit the Decision Maker, because a single object set in condition part that will match one of the decision classes and increase the lower approximation. Both the lower and upper approximations increase the classification accuracy. The higher accuracy rate indicates that the objects belonging to the class have a higher dependency among their condition attributes. Additionally, the number of every attribute value set can influence the class number. If the size of every attribute value set is the same then it is easier to classify the data.

Much of human knowledge is described in natural language. It is very important thing to convert information from computer databases into nature human language. Wikipedia defines natural language (or ordinary language) as language that humans use for general-purpose communication including speech, writing, or sign language. Machine learning techniques are used in natural language processing to deal with uncertain data. The technology used for statistical natural language processing is mainly used for machine learning and data mining, both of which are fields of artificial intelligence that involve learning from data. Natural language processing studies how computer systems understand and generate human language. From a computational perspective, this study proposed the FSBT method, which extends the scope of observations of the rules and converts the results into human natural language to overcome the disadvantages of the RST method.

Tables B.1 and B.2 are listed in Appendix B. The importance of expert knowledge in selecting decision class sets to represent personal investment portfolios is indisputable. The FSBT method works well on special requests or problems, such as exploring consumer behavior, personal portfolios, and sorting problems but must be combined with expert knowledge and incorporated into data analysis. The decision rules generated by the FSBT method are more approximate (rough) and may union classes. However, the coverage of the decision rules (support objects) is increased.

Table 1
Partial decision class as example.

Method	# of objects	Decision rules			
		Class #	Accuracy rate	Support objects	Strength rate (%)
Classical RST	19	1	15/19	15	100
	4	2	4/4	4	100
	2	3	1/2	1	100
	2	4	0/2	0	100
	...				
FSBT	19	1	19/19	49	100
	4	2	4/4	22	100
	2	3	2/2	3	100
	2	4	2/2	2	100
	...				

Table 1 listed the rule quality comparison between the FSBT method and classical RST.

5. Conclusions

We have used the FSBT method to study the asset management market in Taiwan. Based on the survey results, we identified the following types of personal investment portfolio holder:

Conservative: (1) less than 29 years old, (2) college graduate, (3) single, (4) under 4 years work experience, and (5) monthly salary under NT$30,000 (US$900).

Aggressive: (1) less than 39 years old, (2) female, (3) married, (4) a college graduate, (5) 15–19 years work experience, and (6) monthly salary under NT$80,000 (US$2424).

Moderate: (1) less than 29 years old, (2) single, (3) college graduate, (4) less than 9 years work experience, and (5) monthly salary between NT$30,000 (US$900) and NT$80,000 (US$2424).

The proposed method overcomes the problem of too many rules in classical RST that are difficult to implement. FSBT extends classical RST to derive more information so that financial advisors can determine the optimal investment portfolios for their clients. Using the same data sample, financial advisors are better able to prioritize and diversify investments based on their customers' needs.

Acknowledgement

Thanks very much to the reviewers, and Editor, B., Slowinski, for extensive, useful comments which helped to improve the presentation of this article. For the sake of the paper limits to be published, a supplementary material can refer from the web site: http://cid-c0798d7de4c5b827.skydrive.live.com/self.aspx/FSBT%20paper.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2009.03.031.

References

Azibi, R., Vanderpooten, D., 2002. Construction of rule-based assignment models. European Journal of Operational Research 138 (2), 274–293.

Beynon, M., Peel, M., 2001. Variable precision rough set theory and data discretisation: An application to corporate failure prediction. Omega 29 (6), 561–576.

Chan, C., 1998. A rough set approach to attribute generalization in data mining. Journal of Information Sciences 107 (1–4), 169–176.

Chieffe, N., Rakes, G., 1999. An Intelligent system for personal and family financial service. Financial Services Review 8 (4), 261–268.

Dimitras, A.I., Slowinski, R., Susmaga, R., Zopounidis, C., 1999. Business failure prediction using rough sets. European Journal of Operational Research 114 (2), 263–280.

Directorate-General of Budget, Accounting and Statistics, Ministry of the Interior of Executive Yuan, R.O.C. Taiwan. Available from: <http://www.dgbas.gov.tw/mp.asp?mp=1>.

Fan, T., Liu, D., Tzeng, G., 2007. Rough set-based logics for multicriteria decision analysis. European Journal of Operational Research 182 (1), 340–355.

Greco, S., Matarazzo, B., Slowinski, R., 2001. Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129 (1), 1–47.

Hanisch, K., 1994. Reasons people retire and their relations to attitudinal and behavioral correlates in retirement. Journal of vocational behavior 45 (1), 1–16.

Harrison, T., 2003. Understanding the behaviour of financial services consumers: A research agenda. Journal of Financial Service Marketing 8 (1), 6–9.

Hu, Y., Chen, R., Tzeng, G., 2003. Finding fuzzy classification rules using data mining techniques. Pattern Recognition Letters 24 (1–3), 509–519.

Keng, T., Hwa, T., 2004. The role of residential property in personal investment portfolios: The case of MALAYSIA. Pacific Rim Property Research Journal 10 (4), 466–486.

Markowitz, H., 1952. Portfolio selection. Journal of Finance Paper 7 (1), 77–91.

Norland, M., 1988. Japanese household portfolio allocation behavior. Review of Economics and Statistics 70 (1), 135–140.

Ohmae, K., 2006. Lower-middle no Shougeki. Kodansha Ltd. Editorial.

Palma-dos-Reis, A., Zahedi, F., 1999. Designing personalized intelligent financial decision support systems. Decision Support Systems 26 (1), 31–47.

Pawlak, Z., 1982. Rough sets. International Journal of Computer and Information Science 11 (5), 341–356.

Pawlak, Z., 1984. Rough classification. International Journal of Man–Machine Studies 20 (5), 469–483.

Pawlak, Z., 2002. Rough sets, decision algorithms and Bayes' theorem. European Journal of Operational Research 136 (1), 181–189.

Pawlak, Z., 2004. Decision networks. In: Tsumoto, S., Slowinski, R., Komoroski, J., Grzymala-Busse, Jerzy W. (Eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 3066 (1), pp. 1–7.

Plath, D., Stevenson, H., 2005. Financial services consumption behavior across Hispanic American consumers. Journal of Business Research 58 (8), 1089–1099.

Quafafou, M., 2000. α-RST: A generalization of rough set theory. Information Sciences 124 (4), 301–316.

Sawicki, P., Zak, J., 2009. Technical diagnostic of a fleet of vehicles using rough set theory. European Journal of Operational Research 19 (3), 891–903.

Shyng, J., Wang, F., Tzeng, G., Wu, K., 2007. Rough set theory in analyzing the attributes of combination values for the insurance market. Expert System with Application 32 (1), 56–64.

Shyng, J., Tzeng, G., Hsieh, S., Wang, F., 2006. Data mining for multi-domain decision-making based on rough set theory, 2006. IEEE International Conference on Systems, Man, and Cybernetics, Taipei, Taiwan.

Swiniarski, R., Skowron, A., 2003. Rough set methods in feature selection and recognition. Pattern Recognition Letters 24 (6), 833–849.

Teichman, M., Cecconi, P., Bernheim, B., Novarro, N., Monga, M., DaRosa, D., Resnick, M., 2005. How do residents manage personal finances? American Journal of Surgery 189 (2), 134–139.

Hennig-Thurau, T., Klee, A., 1997. The impact of customer satisfaction and relationship quality on customer retention: A critical reassessment and model development. Psychology and Marketing 14 (8), 737–764.

Walczak, B., Massart, D., 1999. Tutorial rough sets theory. Chemometrics and Intelligent Laboratory Systems 47 (1), 1–16.

Witlox, F., Tindemans, H., 2004. The application of rough sets analysis in activity-based modeling, opportunities and constraints. Expert Systems with Application 27 (2), 171–180.

Zadeh, L., 1965. Fuzzy sets. Information and Control 8 (3), 338–353.

Zhai, L., Khoo, L., Fok, S., 2002. Feature extraction using rough set theory and genetic algorithms an application for the simplification of product quality evaluation. Computers and Industrial Engineering 43 (4), 661–676.