



ELSEVIER

Contents lists available at ScienceDirect

Technovation

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

Mining the R&D innovation performance processes for high-tech firms based on rough set theory

Chun-Hsien Wang^{a,*}, Yang-Chieh Chin^b, Gwo-Hshiung Tzeng^{c,d}

^a Department of Bio-industry and Agribusiness Administration, National Chiayi University, Taiwan

^b Department of Management Science, National Chiao Tung University, Taiwan

^c Department of Business and Entrepreneurial Management, Kainan University, Taiwan

^d Institute of Management of Technology, National Chiao Tung University, Taiwan

ARTICLE INFO

Keywords:

R&D innovation
Rule induction
Cause-and-effect relationship
Rough set theory
Flow network graph

ABSTRACT

The research and development (R&D) innovation of firms continues to be viewed as an important source of competitive advantage to academics and practitioners. To explore and extract the R&D innovation decision rules, it is important to understand how the R&D innovation rule-base works. However, many studies have not yet adequately induced and extracted the decision rule of R&D innovation and performance based on the characteristics and components of the original data rather than on post-determination models. The analysis of this study is grounded in the taxonomy of induction-related activities using a rough set theory approach or rule-based decision-making technique to infer R&D innovation decision rules and models linking R&D innovation to sales growth. The rules developed using rough set theory can be directly translated into a path-dependent flow network to infer decision paths and parameters. The flow network graph and cause-and-effect relationship of decision rules are heavily exploited in R&D innovation characteristics. In addition, an empirical case of R&D innovation performance will be illustrated to show that the rough sets model and the flow network graph are useful and efficient tools for building R&D innovation decision rules and providing predictions. We will then illustrate that integrating the flow network graph with rough set theory can fully reflect the characteristics of R&D innovation, and, through the established model, we can obtain a more reasonable result than with artificial influence.

Crown Copyright © 2009 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Research and development (R&D) innovation activity is recognized by its concern with multiple indicators displaying complex structures, uncertainty with many interlocking manufacturing and technological processes, and, consequently, a set of innovation behaviors. Therefore, R&D innovation activity is an important source of competitiveness in many industries. This is particularly true in high-technology industries,¹ characterized by short product life cycles, high uncertainty, and intense competition among new products for market share (Qian and Li, 2003). In

spite of this increasing importance, many high-technology firms have begun to invest heavily in R&D innovation activities to develop novel and innovative products that can help to capture and maintain market share and improve future firm profitability. The R&D effort is a very complex structure with multiple factors to explore to advance R&D resource allocation strategies and translate them into innovations. In response to international competitive pressure, high-technology firm survival and competitive advantage rely upon R&D ability, and hence, innovation in extremely competitive environments (Duysters and Hagedoorn, 2000; Wan et al., 2005). This innovativeness can help capture and maintain market share for improving firm profitability (Wang et al., 2008). There is no doubt about the importance of R&D efforts to high-technology firms as the foundation of their survival. In high-technology industries, where the pace of technical change is speedy, firms place a greater emphasis on R&D efforts toward their products, processes, and technology to overcome technological hurdles and distinguish their offerings from those of competing firms (Thornhill, 2006). R&D efforts can also indicate the innovative competences affecting the performance of firms, particularly in high-technology industries

* Corresponding author: Tel.: +886 5 2732 880; fax: +886 5 2732 874.

E-mail address: chwang@mail.nyu.edu.tw (C.-H. Wang).

¹ In this study, the high-technology industries include electronics, computers, integrated circuits, semiconductors, and telecommunications without covering pharmaceutical firms. In the pharmaceutical industry, as a strongly science-based sector, some of important characteristics such as the long development lead time of drugs and long product life cycles may lead to a difference evaluation results and conclusions. For these reason, the pharmaceutical firms are excluded in our samples.

(Henderson and Cockburn, 1994; Duysters and Hagedoorn, 2001; Hagedoorn and Cloodt, 2003). For these reasons, R&D activity is considered by high-technology firms an important part of maintaining competitiveness and driving profitability.

Many studies rely on empirical studies to analyze R&D innovation activities according to different formal functions, such as the use of the neoclassical production technology function (Griffith et al., 2004), the Cobb–Douglas production function (Tsai and Wang, 2004; Guellec, and Van Pottelsberghe de la Potterie, 2004), panel random effect regressions (Del Monte and Papagni, 2003), OLS regressions (Thornhill, 2006; Coombs and Bierly, 2006), the relev method (Coccia, 2001), and the real options approach (Willigers and Hansen, 2008; Wu and Yen, 2007). Although some of these traditional statistical methods lead to models with a satisfactory ability to analyze R&D innovation activities, they often include a priori assumptions on the data. The advantage that such rough set theory (RST) can be considered applicable here is that it does not rely on other parameters or prior model assumptions. This does not mean that RST does not have any model assumptions. The essence of the RST idea can simply be described by its classes of indiscernible relationships (see Section 3.2). More specifically, RST effectively derives a set of decision rules based on the observed behavior (Beynon and Driffield, 2005). As a result, in this study, we apply RST to mining the decision rule for R&D and innovation to the firm and industry levels. We focus on two firm-level measures of R&D innovation and performance-generating R&D and their innovation decision rules, then examine the causal relationship between R&D innovation activities and, consequently, firm performance. To construct the R&D innovation decision of high-technology firms, we employed the appropriate RST approach, using characteristics and components found to be associated with the R&D innovation decisions and put them forward as inducing decision variables. Empirical analyses of high-technology firms' R&D and innovation decisions are relatively scarce in the existing literature. Therefore, this study is the first attempt to explore the link between R&D innovation and performance decisions of high-technology firms' initial data, or rather any predetermination formal function assumptions.

As can be seen, the R&D innovations of firms have been studied quite extensively for a long period of time. Conventional R&D innovation and performance relationship studies were based on the existing information to build functions or relationship models as a foundation of the measurement decision. Most studies, however, have not appropriately approached how to apply these relationships to traditional production functions, or economic models, before employing these functions, which may not reflect the real relationships in such models. The major problem with most studies measuring R&D innovation and performance is that they rely on predetermined or fitting model measurements (Pawlak, 1991; Pawlak et al., 1995; Pawlak, 1997). This is especially true when researchers adopt a production function or an economic model to measure R&D innovation and performance. Under such a situation, the rule extraction technique is relatively better for inducing a large number of influential attributes and data from irregular, disorderly, and imprecise data, or, more specifically, for discovering relationships in the data. The rules developed by RST are directly translated into the path-dependent flow network graph (Ford and Fulkerson, 1962; Pawlak, 2002; 2004; 2005; Ou Yang et al., 2008) to infer the decision path and parameters between R&D innovation and performance at the firm level in high-technology industries. The use of RST for inducing R&D innovation and performance decisions was advanced as a simple and available way to construct the R&D innovation and performance decisions of high-technology firms. As mentioned previously, this will be the first study to use RST to induce and

explore the relationship between R&D innovation and performance in the high-technology industry. The major objective of this study, therefore, is to better understand the possible rules of specific R&D innovation and performance, by using multiple indicators related to latent variables conditional on the inherent characteristics of the original data. We believe this will provide greater insight for making R&D innovation decisions.

The rest of this article is organized as follows. In Section 2, an overview of the previous relevant work in the domain of R&D innovation is introduced. In Section 3, an RST for inducing the rule of R&D innovation and sale growth ratio is presented. Based on this approach, we define R&D innovation variables, apply their regulation to overall high-technology firm success, and make the link between R&D innovation and percentage sale growth. Next, we design and develop a flow network dependent on the RST decision rule created in Section 4. Finally, conclusions and remarks are proposed in Section 5.

2. Review of prior studies on R&D efforts

To begin our exploration for creating relationships between R&D efforts measures, we first need to explain the literature on the characteristics of R&D innovation and performance to construct some reference points and rules. Naturally, R&D efforts can accelerate the accumulation of knowledge and technological strength. It can also help determine a firm's performance. In the current literature, there exist many studies proposing various methods for establishing and identifying the relationships between R&D innovations and firm performance. Earlier empirical studies have examined associating R&D expenditures strongly with sales growth (Morbey and Reithner, 1990), profit and productivity (Baumol and Wolff, 1983; Dosi, 1988; Morbey and Reithner, 1990). In light of this logic, Franko (1989) argued that R&D investments were positively related to long-term performance. Ettl (1998) examined the effect of R&D intensity on manufacturing performance. He found that R&D intensity has significantly contributed to the increasing market share and improvements in manufacturing agility. Wakelin (2001) applied the Cobb–Douglas function to estimate United Kingdom (UK) manufacturing firms' R&D expenditures, finding these expenditures to have a significant positive impact on productivity growth. Brown and Gobeli (1992) applied the results of a case study to develop a system for measuring R&D productivity. They developed a conceptual model and determined the top 10 qualitative and quantitative indicators of R&D productivity measurements. In line with this is the conclusion by Werner and Souder (1997) that R&D performance evaluation is an integrated measurement system in which qualitative and quantitative metrics may be combined and used simultaneously. Dressler et al. (1999) suggested that the cost-saving ratio was an appropriate approach used to measure R&D performance. In addition to the papers mentioned previously, a few other empirical findings suggested that R&D has a positive impact on a firm's performance (Ito and Pucik, 1993; Long and Ravenscraft, 1993; Lee and Shin, 1995). In summary, a majority of the literature has empirically demonstrated the significant impact of R&D innovation activities on the performance of firms; however, these studies did not examine whether the original data as appropriate for the exploration and connections between R&D innovation and performance in the empirical model. So far, evidence permitting inference from the existing information is quite limited.

Firms are concerned about R&D resources deployment and productivity because investing in R&D efforts has become an important innovation source for their production activities and strategic decisions. Several economic studies have attempted to

estimate the relationship between R&D and the growth of productivity at the firm level. Baumol and Wolff (1983) illustrated that the scale of R&D activity directly translated to the rate of growth of productivity in manufacturing, which, in turn, affected the relative cost of R&D and, hence, its demand. Guellec and Van Pottelsberghe de la Potterie (2004) argue that R&D activities are important source of knowledge on multifactor productivity growth, a significant determinant of long-term productivity growth. In the same vein, Frascati Manual (OECD, 1993) defined R&D efforts as being comprised of creative work undertaken on a systematic basis to increase the stock of knowledge. The purpose of this stock of knowledge is to devise new applications (OECD, 1993). Smith et al. (2004) used empirical evidence to examine the link between investment in R&D and productivity using data for Danish manufacturing firms. The influence of factors such as ownership, innovative characteristics and sources of funding were used, but none of the factors were found to have an impact on firm productivity. On the contrary, Hall and Mairesse (1995) used the same approach to explore the relationship between R&D and productivity in French manufacturing firms. Their major finding was that firms had a longer history of R&D expenditures, which improves the quality of the R&D elasticity estimates. They also found that the relationship between R&D capital and productivity for French manufacturing firms was positive. Thornhill (2006) found evidence that the high-technology manufacturing sector contains a higher percentage of firms introducing national or world-first new products than the low-technology sector does. This indicates that the high-technology industry, with greater aggregate levels of R&D intensity, is home to higher rates of firm-level innovation activity. In his empirical study, Thornhill confirmed that innovation was positively associated with firm performance. A similar study was undertaken by Wang and Tsai (2003), who estimated the impact of R&D on productivity in the high-technology and conventional firms using total factor productivity at the firm level. Their empirical study found that the R&D investment had a particularly significant impact on firm productivity growth; the R&D output elasticity in high-technology firms was significantly greater than that of conventional firms. These studies suggest that firms' increased reliance on R&D efforts could affect the growth of the productivity of firms, and consequently, following innovation activities.

All of the previously mentioned studies recognize R&D efforts as having a direct or indirect impact on the performance, or productivity, of a firm. These models make strong assumptions to permit for a causal link between observed independent explanatory and dependent (explained) variables. However, this need not be true regarding these assumptions. One possible reason could be that these studies were based on a predetermination assumption of a direct relationship between R&D and performance within the measurement model, rather than on a logical reasoning approach based on original data. In fact, these measurement models were dependent on casual observation, implicit goals, intuitive norms, and pre-artificial subjective judgments to make their predictions. The RST inductive approach proposed in this study is an attempt to overcome some of the shortcoming of previous models and to induce, extract, and measure, the rules and regulations related to R&D innovation and performance.

The RST approach is an appropriate technique that can be used to solve these disadvantages. The RST approach is recognized by its dependence on the original data, structured reasoning by latent rules, controlled comparisons and similarity-based reasoning. According to Pawlak et al. (1995), the RST is a useful method for use in the data reduction (elimination of superfluous data), discovery of data dependencies, estimation of data significance, generation of decision (control) algorithms from the data, approximate classification of data, discovery of similarities or

differences in data, discovery of patterns in data, and discovery of cause-effect relationships. Therefore, the RST provides a useful and efficient way to extract, induct, classify, and discover hidden information, based on large-scale empirical data without requiring a probability distribution. Compared to other approaches for handling inconsistent information, the use of the RST carries many advantages (Pawlak, 1996, 1997). It does not require any preliminary or additional information about the initial data. In addition, it is more appropriate than standard statistical methods when the empirical or experimental data is too small (Pawlak, 1991). The RST was selected for our study because it permits us to identify important decision rules and cause-effect relationships of R&D innovation efforts. In high-technology firms' R&D innovation efforts, it was necessary to know which aspects of R&D statements were needed to decide the future of their R&D innovation. More specifically, the RST led to the creation of rules linking the dependent to the independent variables. This could be valuable for the analysis of R&D innovation activities and decisions.

RST draws conclusions from the initial data without referring to prior and posterior given functions when the data becomes available. The R&D innovation decision rule inference, based on the RST, is used to verify prior information and extract knowledge when the initial data becomes available, whereas the RST is based on the decision rule inference in the initial data.

3. Rough set theory and flow network graph algorithm

The original rough set theory (RST) was proposed by Pawlak (1982), who drew upon set theory in its creation. RST is an effective mathematical approach for discovering hidden deterministic rules and associative patterns in all types of data and for handling unknown data distributions and information uncertainty. RST is especially useful for analyzing imprecision, uncertainty, or vagueness in the classification of objects in a set. RST can be used to deal with quantitative and qualitative attributes simultaneously without requiring any a priori information about the probability distribution of the data. Therefore, many different studies have adopted the RST approach to extract rule and patterns from original data and unclassified information. RST has been widely applied in various domains, such as knowledge acquisition, knowledge discovery from initial databases, expert systems, decision analysis and rule induction, and inductive reasoning from original data. To understand the initiatives at building decision rules for R&D innovation in the high-technology industry, we may start by using RST approach to extract and to discover the R&D innovation knowledge what we need.

3.1. Basic concepts of rough set

According to Pawlak (1982), the basic operation of RST may be described as an information table or decision table, which can be represented by a set of objects dependent on multi-valued attributes represented. An information system can be represented by the quintuple $S=(U,Q,V,f)$, where U is a finite and non-empty set of objects called the universe composed of a certain set of objects, and Q is a non-empty finite set of attributes describing objects in the universe. It can be divided into two subsets C and D , which denote the finite set of condition attributes and the set of decision attributes, respectively. $V = \cup_{a \in Q} V_a$, in which V_a is the set of attributes for values Q , and $f:U \times Q \rightarrow V$ is an information function such that $f(x,a) \in V_a$ for every $a \in Q$ and $x \in U$ (Pawlak, 1991; Fan et al., 2007). For any application of RST, objects can be interpreted as cases, states, processes, patients and observations. Attributes can be interpreted as features, variables and characteristic conditions (Tay and Shen, 2002; Shyng et al., 2007).

3.2. Indiscernibility

A central premise of rough set philosophy is that the knowledge consists in the ability to classify. The most important approach in knowledge classification is indiscernibility, which can be considered a formal mathematical tool for extracting and discovering facts from imperfect data. The indiscernibility relation generated in the universe of discourse is the mathematical basis for the rough set theory. Discretization can convert continuous attributes into discretized ones and can synchronously remove redundant and irrelative attributes. Any set of all indiscernible objects is called an elementary set and forms a basic granule of knowledge about the universe. In the information table $IS=(U, Q, V, f)$, assume that $Q=C \cup D$ and $C \cap D = \emptyset$, where C is a set of condition attributes, and D is a set of decision attributes. Let $P \subseteq Q$, which generates an indiscernibility relation $IND(P)$ on U , called the P -indiscernibility relation. Obviously, $IND(P)$ is an equivalence relation for any P . The indiscernibility relation $IND(P)$ is defined as follows:

$$IND(P) = \{(x, y) \in U^2 \mid f(x, a) = f(y, a), \forall a \in P\} \quad (1)$$

$U/IND(P) = \{C_1, C_2, \dots, C_k\}$ is a partition of U by P , and every C_i is an equivalence class. For $\forall x \in U$, the equivalence class of x in relation $U/IND(P)$ can be defined as follows:

$$[x]_{IND(P)} = \{y \in U \mid f(y, a) = f(x, a), \forall a \in P\} \quad (2)$$

3.3. Approximations of sets

A rough set-based rule induction technique can be expressed by a pair of crisp sets called the lower and the upper approximation. The lower approximation contains all objects that certainly belong to that category. The upper approximation consists of all objects that can possibly belong to that category. A rough set offers a means to describe vague classes through these lower and upper approximations. In the information table $IS=(U, Q, V, f)$ assume that $P \subseteq Q$ and $Y \subseteq U$. Then, the lower approximation of $P(X)$ contains all the objects in the universe U that can be certainly classified as belonging to class X by the set of attributes, P , such that

$$\underline{P}(X) = \cup \{Y \in U/IND(P) \mid Y \subseteq X\} \quad (3)$$

The upper approximation of $\underline{P}(X)$ contains all the objects in U that can be possibly classified as belonging to class X by the set of attributes, P , such that

$$\overline{P}(X) = \{Y \in U/IND(P) \mid Y \cap X \neq \emptyset\} \quad (4)$$

The P -boundary region of set X , denoted by $BN_p(X)$, is defined as

$$BN_p(X) = \overline{P}(X) - \underline{P}(X) \quad (5)$$

The set $BN_p(X)$ is the set of elements that cannot be certainly classified to X using the set of attributes P .

With every set $Y \subseteq U$ we can estimate the accuracy of approximation of the set Y by P using

$$\alpha_P(X) = \frac{\text{card}(\underline{P}(X))}{\text{card}(\overline{P}(X))} \quad (6)$$

where the cardinality of a set "card" is the number of objects contained in the lower (upper) approximation of the set X . Clearly, $0 \leq \alpha_P(X) \leq 1$ for every non-empty $P \subseteq Q$ and $X \subseteq U$. If X is definable in U , then $\alpha_P(X) = 1$, and if X is undefinable in U then $\alpha_P(X) < 1$. Moreover, the quality of approximation of classification X by P (or quality of classification in short) is defined

as follows:

$$\alpha_P(X) = \frac{\sum_{i=1}^n \text{card}P(X)}{\text{card}(U)} \quad (7)$$

The coefficient of $\rho_P(X)$ is called the quality of approximation of classification X by the set of attributes P or, in short, the quality of classification. It indicates the ratio of all P -correctly classified objects to all objects in the system.

3.4. Attributes reductions and core

In the information table, some attributes may be redundant due to their being irrelative or unnecessary and thus can be eliminated without losing essential classificatory information. To do so, the main computational efforts in the processing of data in RST are associated with attribute reduction and core. An attribute reduction is the minimal subset of attributes that provides the same quality of classification as the set of all attributes. The means of the core is the common part (intersection) of all reductions. Given A and $B \subseteq Q$, a reduct is a minimal set of attributes such that $IND(A) = IND(B)$. Therefore, a reduct is a minimal non-redundant set of attributes that provides the same quality of classification as the set of all attributes. Let $RED(A)$ be the set of all reducts for A . The intersection of all reducts $CORE(A) = \cap RED(A)$ is referred to as a core of A . The core is a collection of the most important attributes in the decision table.

3.5. Decision rules extraction

A set of condition attributes C and decision attributes D can be derived according to the decision table. The decision table S is a deterministic or exact decision rule if $C \rightarrow D$. Otherwise, the rule is non-deterministic or approximate. The deterministic decision table uniquely expresses the decisions to be defined when particular conditions are satisfied. On the other hand, elements of the non-deterministic decision table are not uniquely determined by the conditions. Therefore, a set of decision rules can be derived from a decision table for decision analysis. The procedure of capturing decision rules from a set of initial data is known as induction (Pawlak, 1991). An induced decision rule can be expressed as a logical manner.

IF a conjunction of elementary conditions; **THEN** a disjunction of elementary decisions

Simply, a typical form of decision rule can be expressed as **IF**_condition(s)_**THEN**_decision(s). The rules are logical statements "if...then..." relating the condition and decision classes. The decision rule reflects a relationship between a set of conditions and a conclusion or a decision. Mark and Munakata (2002) argue that the extract rules using rough sets is relatively simple and straightforward and that no extra computational procedures are required before rules can be extracted. Therefore, in this study, construction of the decision rules is performed based on upper and lower approximations extracted from the decision table.

3.6. The causal-and-effect of decision rules based on flow network graph

Under the assumption of decision rules of the R&D innovation characteristics, we find a causal-and-effect path-dependent figure that depends on the rule and initial characteristics of R&D innovation potential. The flow graph, proposed by Ford and Fulkerson (1962), is a powerful tool for explaining a path-dependent relationship based on the rough sets of decision rules. Branches of the flow network graph are interpreted as decision

rules, whereas the flow graph is supposed to describe a decision algorithm. According to the flow graph and Bayes' theorem (Pawlak, 2002), the model was used to capture and describe the nature of decision processes within flow network graphs rather than description of flow optimization. The relationship between flow network graphs and decision algorithms is presented as follows (Pawlak, 2004, 2005; Ou Yang et al., 2008).

A flow graph is a directed acyclic finite graph $G=(V,\beta,h)$ where V is a set of nodes, $\beta \subseteq V^2$ is a set of directed branches, $h:\beta \rightarrow R^+$ is a flow function and R^+ is the set of non-negative real numbers. A branch $(x,y) \in \beta$, then x is an input of y and y is an output of x . The throughflow of a branch $(x,y) \in \beta$ and can be defined as $r(x,y)$. For $(x,y) \in \beta$ then $r(x,y)$ is a throughflow from x to y . The input of a node $x \in V$ is the set $I(x)=\{y \in V | (y,x) \in \beta\}$, and the output of a node $x \in V$ is defined as $O(x)=\{y \in V | (x,y) \in \beta\}$. Based on these concepts, the input and output of a graph G are defined as $I(G)=\{x \in V | I(x) \neq \emptyset\}$ and $O(G)=\{x \in V | O(x) \neq \emptyset\}$. For every node x in flow graph, inflow is defined as $h_+(y)=\sum_{x \in I(y)} h(x,y)$ and outflow is defined as $h_-(y)=\sum_{y \in O(x)} h(x,y)$. Similarly, the inflow and outflow of the whole flow graph can be defined as $h_+(G)=\sum_{x \in I(G)} h_-(x)$ and $h_-(G)=\sum_{x \in O(G)} h_+(x)$, respectively. We assume that for any node x in a flow graph G , $h_+(x)=h_-(x)=h(x)$. In a similar way, a throughflow of the whole flow graph G is expressed as $h_+(G)=h_-(G)=h(G)$.

To measure strength of every branch (x,y) in a flow graph $G=(V,\beta,h)$, we define the strength $\rho(x,y)=h(x,y)/r(G)$. Obviously, $0 \leq \rho(x,y) \leq 1$. The strength of the branch simply expresses the amount of total flow through the branch. Every branch (x,y) of a flow graph G is associated with certainty and coverage coefficients. The certainty and coverage of every branch are defined as $\text{cer}(x,y)=\rho(x,y)/\rho(x)$, and the $\text{cov}(x,y)=\rho(x,y)/\rho(y)$, respectively, where $\rho(x,y)=h(x,y)/h(G)$, $\rho(x)=h(x)/h(G)$ and $\rho(y)=h(y)/h(G)$ are normalized throughflow, and $\rho(x) \neq 0$, $\rho(y) \neq 0$, and $0 \leq \rho(x,y) \leq 1$. The meaning of certainty coefficient expresses outflow distribution between outputs of a node, whereas the coverage coefficient exhibits how inflow is distributed between inputs of the node. The above coefficients simply explain some properties of flow distribution among branches in the whole flow network graph. The basis of the flow graph theory can be traced back to Ford and Fulkerson (1962). More advanced topics of decision and flow networks are discussed in Pawlak (2002, 2004, 2005) and Ou Yang et al. (2008). Hence, the flow network graph is basically meant as a powerful tool for modeling flow information represented by a set of decision rules. This is why we decided to adopt the flow network graph as tool to present the decision processes in our research. Hence, in this study also applied influence diagrams to help the decision makers or managers by through a set of the R&D innovation decision rules to the description of the appropriate decision paths and directions. The primary reason that such an approach can be considered employable here is that it effectively derives a set of related rules based on the results of RST. The influence diagram tries to connect as much rules as possible from the contextual aspects of the data, hence the relationship between RST and influence diagram is quite complementary. This integration model can serve as a valuable indicator of the direction into which possible further analysis can explore more hidden information what we need.

4. An empirical case of high-technology firms

Many authors have attempted to define what the high-technology industry is. A broad definition of a high-technology firm is one whose business activities are heavily dependent upon innovation in science and technology approaches (Medcof, 1999). High-technology industries are also defined as those industries

investing proportionally more heavily in scientific and technological activities than other industries (Butchart, 1987). The more commonly accepted approach is to define high-technology firms on the degree of expenditure on R&D as a percentage of sales greater than 5% (Balkin et al., 2000). According to Reebbe (1990), a high-technology industry includes activities in which rapid technological change and high inputs of scientific R&D expenditure and employment are producing new innovative and technologically advanced products. Diaz and Gomez-Mejia (1997) and Keeble and Wilkinson (2000) argued that the primary characteristics of high-technology firms are high levels of R&D intensity and high levels of radical innovation activities. Therefore, a major characteristic of high-technology firms is their proportion of their resource investment in R&D innovation activities. We adopted a broad spectrum of sectors to define high-technology industries, such as telecommunications, computer hardware and software, semiconductors, electronics, biotechnology, medical and pharmaceuticals.

Several authors observed that the development of the high-tech industry in industrializing countries as frequently taking low capital, land, or labor-intensive manufacturing technology represents a major source of competitive advantage (Porter, 1990; Chang and Hsu (1998). Many of Taiwanese high-tech firms achieve low-cost production through mass production of standardized products. Such production model provide local firms moved quickly to accumulate experience and manufacturing capabilities, thus Taiwanese high-tech firms cultivated more advanced production techniques and knowledge to upgrade their R&D and innovativeness capabilities. These developments make it is imperative for R&D managers working for Taiwanese high-tech firms to upgrade their R&D innovation capabilities and increase level of local competitiveness. Competition drives the innovation process and the upgrading of capabilities (Porter, 1980, 1990); that is, more innovative and with better access to the resources necessary for their productive activities is need. Consistent with the Porters' perspective (Porter, 1990), firm competitive advantages are created by the sustainable low-cost or differentiated position against competitors.

In the past two decades, Taiwan high-technology industries have often achieved a dominant market position based on their superior R&D and innovation capabilities. The Taiwanese high-technology industry is well known for its research, design, development, innovation, and manufacturing capabilities. The primary reasons can successfully ascribe contents to high-technology industries, including foreign computer manufacturer investment the government recognizing the development of high-technology industries for maintaining economic growth (Chen and Huang, 2004), and government participation and intervention in the development high-technology science parks. Reflecting this process, the Taiwan high-technology industries have continually invested in R&D innovation activities to upgrade their levels of technology development. As a result, many firms have abandoned their original equipment manufacturing (OEM) model and became involved in design or brand manufacturing (ODM or OBM, respectively). The successful transition from OEM to ODM or OBM models within the Taiwan high-technology industry is because of high R&D intensity and well-defined innovation activities permitting firms to upgrade their related production technology and knowledge to sustain their competitiveness. This is also why many high-technology firms engage in developing and producing advanced technology and improving efficient manufacturing operations to carry out and speed up their innovation process. As a consequence, high-technology firms undertaking R&D efforts are able to accelerate the speed of the innovation of products, services, and technical competence to enhance their performance and competitiveness. In this section, we adopted the

RST approach which includes: (1) selecting variables and data; (2) calculating the approximation; (3) finding the reductions and core attributes; and (4) incorporating decision rules into a flow network graph as the final decision algorithm. The results are used to predict R&D innovation efforts and strategic implementation in the high-technology industry.

4.1. Selection variables and data

To increase the accuracy of decision variables, we invited three scholars from a university technology management department and three managers from high-technology industry with an average of over 6 years' of experience in R&D management and technological innovation management. The participants were asked to identify whether the selected variables can be viewed as R&D innovation criteria. Based on the literature review and on our experts' opinions, nine criteria for R&D innovation relationships were identified for extracting R&D innovation rules of firms. We have employed qualitative and quantitative variables. This is one of the main advantages of the RST approach's using two kinds of variables. To induce the rules and regulations of R&D innovation and performance, there are eight conditional attribute variables and one decisional attribution variable employed in the logical reasoning analysis of the RSTs. The decision attribution variable in the firm performance criterion is sales growth ratio, computed as the average percentage change in revenue over the period of 2002–2006. The sales growth ratio was selected as the decision criterion, because it should directly reflect the impact of introducing new products (Thornhill, 2006) and can be viewed as a performance measurement index of the firms (Del Monte and Papagni, 2003; Lee and Shin, 1995). Sales growth was used to measure R&D output, because of the characteristics of a short product life cycle and intense competition in the high-technology industry. More specifically, the successful transition from OEM to ODM or OBM models within the Taiwan high-technology industry is because of high R&D intensity and well-defined innovation activities permitting firms to upgrade their related production technology and knowledge to sustain their competitiveness. These characteristics may accelerate high-technology firms' R&D activities and shorten product manufacturing time, and hence, reflect quickly on the sales growth in world markets. In this study, the dependent variable is the sales growth ratio, identified as in Grupp and Maital (2000) and Koschatzky et al. (2001). The independent variables cover the relationship theory based on the determinants of R&D innovation activities. We will provide a detailed definition of variables and sources in the following:

- *Total assets*: total assets are the typical inputs in R&D activities.
- *R&D expenditures ratio*: The R&D expenditure ratio is measured by the ratio of R&D expenditures to total sales. R&D expenditure serves as the indicator of input or firm innovativeness.
- *Proportion of R&D researchers*: R&D researchers are an important part of motivating and engaging in R&D innovation activities. That is, R&D researchers are directly and deliberately engaged in productivity-enhancing and value-enhancing activities.
- *Number of patents*: The number of patents can be viewed as an indicator of R&D success (Graves and Langowitz, 1996; Kim and Oh, 2002; Wan et al., 2005). The number of patents also measures the volume of firm research activities and the impact of a firm's research on subsequent innovations.
- *The proportion of exports*: Exports indicate the higher rate of exports for firms with high R&D innovation input. This finding is consistent with the findings of Sterlacchini (1999) and Roper

and Love (2002), that R&D intensity increases either the possibility of innovative products' being an exporter or the share of exports represented by sales.

- *Sales growth ratio*: Sales revenue represents the profitability associated with R&D and innovation activities' resulting in new products and services. An important indicator of the realization of product innovations is the share of new products in sales revenue (Koschatzky et al., 2001). Grupp and Maital (2000) examined the R&D innovation activities of the largest Israeli firms and found an association between perceived innovativeness and significant increases in sales revenues and intended future profitability. Thornhill (2006) also noted that firms that undertake innovation activities are likely to enjoy revenue growth.
- *Located in a science park*: High-technology firms located in science parks may impact their R&D innovation capabilities. The well-developed science park can provide many benefits to high-technology firms, such as shared local markets and resources, knowledge spillover effects, and low coordination costs.
- *Overseas branch*: Building an overseas branch is one way to acquire R&D and innovation information, advanced techniques, and knowledge from an advanced country.
- *Return on investment (ROI)*: ROI is the easiest financial criteria to calculate for R&D outcomes, which are relatively stable and predictable. It is also the single most important indicator of R&D innovation performance (Hartmann, et al., 2006; Walwyn, 2007).

This study is based on panel data gathered from multiple databases related to the high-technology industry in Taiwan. The study sample was comprised of all of the over-the-counter (OTC) high-technology industry firms listed in the Taiwan Stock Exchange Corporation (TSEC) database for which continuous financial data was available for the period from 2002 to 2006. Follow-up information was obtained from multiple databases to increase accuracy and reduce the error rate. A final sample of 181 high-technology firms was thus obtained. The sample included firms involved in electronics, computers, integrated circuits, semiconductors, and telecommunications.

4.2. The rules-based prediction of R&D innovation characteristics

Using the R&D innovation variables of the high-technology firms, nine attributes were available. These data were pre-processed to construct the information table, which represents knowledge in a RSTs model. The information table contains eight conditional attributes and one decision attribute presented in Table 1.

In the RST, the most important approach in the ability of classification is indiscernability, a formal mathematical tool for extracting and discovering facts from imperfect data. Discretization can convert continuous attributes into discretized ones and synchronously remove redundant and irrelative attributes. There is no general way to define the optimal boundary values. The use of experts' opinions, according to their experience and knowledge, is the best way to identify a set of decision problems (Dimitras et al., 1999). Therefore, we asked experts to discretize the continuous R&D innovation indicators, providing norms according to their professional knowledge and experiences. The intervals were determined by experts based upon the average and standard deviation from the initial data. The interval peer assessment was conducted in three rounds during the RST mining processes. These experts had similar perspectives in evaluating these intervals, indicating that experts' options showed adequate reliability of

Table 1
R&D innovation attributes description.

Attributes/variables in the information table	Descriptions
a_1 Located in a science park	A corporation is located in the science park. A science park is a specific land for related industries and organizations to locate together.
a_2 Overseas branch	A corporation builds its subsidiary abroad in order to acquire host resources and absorb other firms' knowledge, technique, and experiences.
a_3 Return of investment (ROI)	Return of investment ratio is the ratio of a corporation's gains or losses on an investment relative to the amount of money invested.
a_4 Number of patents	The number of approved patent applications of each firm is used to measure innovation capability.
a_5 Proportion of export	The proportion of export of a firm is total export volume divided by total market sales.
a_6 Proportion of R&D researchers	Proportion of R&D researchers to total employees is the employee with research and development ability to the total amount of a corporation's employees.
a_7 R&D expenditures ratio	The R&D expenditure ratio is measured by the ratio of R&D expenditures to total sales.
a_8 Total assets	The amount of cash and property of a corporation.
d_1 Sales growth ratio (average past three years sales growth)	The sales growth ratio is an important performance indicator.

Table 2
Sub-intervals and their codes for conditional attributes.

Attributes	Intervals/number of intervals		
	1	2	3
a_1	Yes	No	
a_2	Yes	No	
a_3	$(-\infty, 2.96)$	$[2.96, 15.96)$	$[15.96, +\infty)$
a_4	$(0, 11.5)$	$[11.5, +\infty)$	
a_5	$(0, 0.58)$	$[0.58, 0.87)$	$[0.87, +\infty)$
a_6	$(0, 0.17)$	$[0.17, +\infty)$	
a_7	$(0, 3.63)$	$[3.63, +\infty)$	
a_8	$(0, 120,787)$	$[120,787, 238,593)$	$[238, 593, +\infty)$

Note: The symbols “[” and “]” indicate the left and right closed boundaries, and “(” and “)” denote the left and right open boundaries, respectively.

assessing indiscernibility design. The primary purpose of discretization is the prohibition of drawing general conclusions from data in terms of dependencies, reducts and decision rules (Dimitras et al., 1999). When firms exhibit the values of these attributes in the same intervals, this implies that they have very similar R&D innovation characteristics and behaviors. The intervals proposed for the discretization are presented in Table 2. The codes used to refer to each sub-interval do not represent any preferential ordering. Since the RST is based on the indiscernability relationship, one may consider any order. The coded information table prepared for the future analysis consisted of 181 high-technology firms according to nine attributes.

Based on the decision rules extraction procedures of the RST, a large number of R&D innovation rules can be generated. These were important for knowing whether all rules played an important role in the classification process. The first results obtained from RSTs analysis of the coded information table were the approximations of the decision classes and the quality of their classifications. These results revealed that the data were very well

categorized and appropriate for predicting R&D innovation efforts. As we can see in Table 3, the accuracy of the classification in terms of positive relationships with sales growth ratio was 100% and in the negative relationship was also 100%. In this manner, the results indicate that eight conditional attributes play important roles in determining the sales growth ratio and are appropriate for predicting R&D innovation characteristics, whether a firm has a positive or negative sales growth ratio. In Table 3, we can identify the quality of the classification as 100%, meaning that all samples were correctly classified.

To increase the classification rate and acquire the reduct attribute sets during the RSTs analysis processes, the reduction of conditional attributes through an exhaustive algorithm is employed to determine the superfluous attributes. We have obtained the result of the value of the positive region of reduct as 1.0, thus there are no superfluous attributes in our analysis. Therefore, the core and reduced set consists of eight attributes $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}$, representing all relevant attributes in the table. This result shows the importance of these eight variables for forecasting R&D innovation behaviors in the high-technology firms.

Through RSTs analysis, 204 rules were obtained from the coded informational table (cf. Section 3.4). The table illustrates 125 rules in positive relationships and 79 rules in negative relationships with the sales growth ratios of high-technology firms, respectively. To interpret the rules, we set up the threshold value of the percentage of training data as 10% for each decision class after consulting with experts; thus, we only considered 16 rules (10 for class positive relationships with sales growth ratios and the other ones for class negative). These rules have been selected with attention to categorization in terms of correctly classified high-technology firms, as well as in terms of R&D innovation characteristics and its behavioral understanding. This means over 10% of the samples with three or more of the same conditional attributes could be classified with the same decision attributes.

Focusing on the role of the R&D innovation in determining performances, we further observed the effects of R&D innovation decision rules on the sales growth ratio, as illustrated in Table 4. The frequent occurrences of the variables in the decision rules table include total assets (11 times), proportion of exports (10 times), proportion of R&D researchers (8 times), located in a science park (7 times), R&D expenditures ratio (6 times), ROI, (5 times), overseas branch (4 times), and number of patents (3 times). Therefore, we can see that, in Table 4, some of the variables had a higher degree of dependence associated with R&D innovation activities, which may impact the success or failure of R&D innovation efforts. These results illustrate the different degrees of importance of variables for forecasting R&D success, which could help manage firms develop R&D strategies.

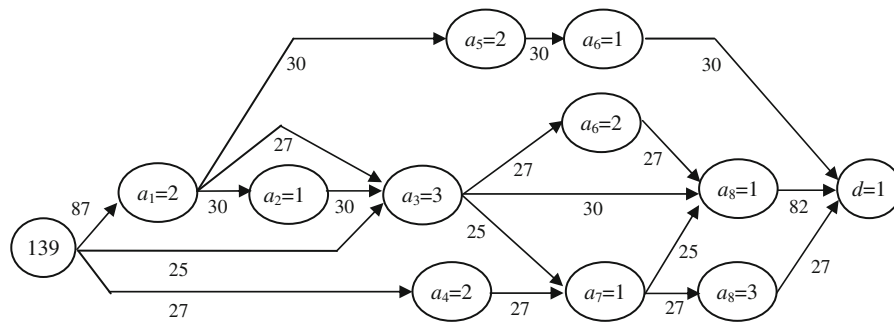
As illustrated by our results, the decision rules generate two directions of sales growth ratio of high-technology firms. One is that the R&D innovation has highly significant positive effects on the sales growth ratio. The most important of the decision rules of the R&D innovation characteristics of the high-technology firms are the following: a firm (1) is not located in a science park, (2) owns an overseas branch, (3) the ROI ratio is greater than 2.96%, (4) has patents that number over 11.5, and (5) has an R&D expenditures ratio less than 3.63%. Another result is that the R&D innovation characteristics have negative effects on the sales growth ratio. More specifically: (1) a high-technology firm is not located in the science park, but owns an overseas branch, (2) the proportion of exports is greater than 87%, and (3) the amount of a firm's total assets is between NT\$ 120,787 (in thousands) and NT\$ 238,593 (in thousands).

Table 3
Accuracy of classification and quality of classification.

Sales growth ratio	Numbers of high-tech firms	Lower approximation	upper approximation	Accuracy of classification	Quality of classification
Positive	158	158	158	1.00	1.00
Negative	23	23	23	1.00	

Table 4
R&D innovation with sales growth ratio decision rules.

Rules	Number of matching firms	% of training data covered by the rule (%)
<i>Positive sales growth ratio</i>		
1 IF ($a_1=2$) & ($a_2=1$) & ($a_3=3$) & ($a_8=1$) THEN (Positive sales growth ratio)	30	18.99
2 IF ($a_1=2$) & ($a_5=2$) & ($a_6=1$) THEN (Positive sales growth ratio)	30	18.99
3 IF ($a_4=2$) & ($a_7=1$) & ($a_8=3$) THEN (Positive sales growth ratio)	27	17.09
4 IF ($a_1=2$) & ($a_3=3$) & ($a_6=2$) & ($a_8=1$) THEN (Positive sales growth ratio)	27	17.09
5 IF ($a_3=3$) & ($a_7=1$) & ($a_8=1$) THEN (Positive sales growth ratio)	25	15.82
6 IF ($a_1=2$) & ($a_5=3$) & ($a_6=2$) THEN (Positive sales growth ratio)	23	14.56
7 IF ($a_3=2$) & ($a_7=1$) & ($a_8=3$) THEN (Positive sales growth ratio)	21	13.29
8 IF ($a_1=2$) & ($a_3=3$) & ($a_5=1$) THEN (Positive sales growth ratio)	19	12.03
9 IF ($a_5=2$) & ($a_6=1$) & ($a_8=1$) THEN (Positive sales growth ratio)	18	11.39
10 IF ($a_5=3$) & ($a_6=2$) & ($a_8=2$) THEN (Positive sales growth ratio)	18	11.39
<i>Negative sales growth ratio</i>		
1 IF ($a_5=3$) & ($a_6=2$) & ($a_8=2$) THEN (Negative sales growth ratio)	5	21.74
2 IF ($a_1=2$) & ($a_5=3$) & ($a_8=2$) THEN (Negative sales growth ratio)	4	14.29
3 IF ($a_2=2$) & ($a_5=3$) & ($a_8=2$) THEN (Negative sales growth ratio)	3	13.04
4 IF ($a_2=2$) & ($a_4=2$) & ($a_5=3$) & ($a_6=2$) & ($a_7=2$) THEN (Negative sales growth ratio)	3	13.04
5 IF ($a_1=2$) & ($a_2=1$) & ($a_4=2$) & ($a_6=1$) & ($a_7=2$) THEN (Negative sales growth ratio)	3	13.04
6 IF ($a_5=3$) & ($a_7=2$) & ($a_8=2$) THEN (Negative sales growth ratio)	3	13.04



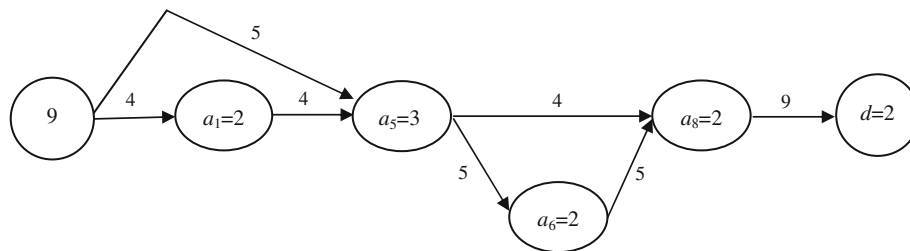
$a_1=2$ means that a high-tech firm does not establish itself in a science park; $a_2=1$ means that the high-tech firm does not build an overseas branch; $a_3=3$ means that the return on a high-tech firm’s investment is above 15.96; $a_4=2$ means that the number of patents of high-tech firm is above 11.5; $a_5=2$ means that the proportion of exports of a high-tech firm ranges from 0.58 to 0.87; $a_6=1$ means that the proportion of R&D researchers in a high-tech firm is below 0.17; $a_6=2$ means that the proportion of R&D researchers in a high-tech firm is above 0.17; $a_7=1$ means that the R&D expenditure ratio of high-tech firm is below 3.63; $a_8=3$ means that the total assets of a high-tech firm is above NT\$238,593; $d=1$ means that the sale growth ratio of a high-tech firm is positive.

Fig. 1. Decision flow graph and rule-set of positive sale growth ratio.

4.3. The causal-and-effect flow network graph

To denote the causal and effect relationship of entities, they must be perceived as real observed variables. Figs. 1 and 2 represent the relationships among the R&D innovation variables. The decision flow network graph in this study was used to represent the relationship between R&D innovation and performance and to provide information for further model

refinement. A causal loop diagram provides a bridge that consists of variables connected by arrows expressing the causal-and-effect relationships among the R&D innovation variables. In a similar vein, Brown and Gobel (1992) proposed that the productivity measurement of R&D provides a clear cause and effect relationship for understanding and measuring the relative R&D innovation activities. Therefore, the rational R&D innovation strategy is to select a better set, since the anomaly affects each



$a_1=2$ means that a high-tech firm does not establish itself in a science park; $a_5=3$ means that the proportion of exports in a high-tech firm is above 0.87; $a_6=2$ means that the proportion of R&D researchers in a high-tech firm is above 0.17; $a_8=2$ means the total assets in a high-tech firm ranges from NT\$120,787 to NT\$238,593; $d=2$ means the sale growth ratio in a high-tech firm is negative.

Fig. 2. Decision flow graph and rule-set of negative sale growth ratio.

Table 5
Decision rules of positive sales growth ratio with minimum support value ≥ 18 .

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d_1	Support	Certainty	Strength	Coverage
2	1	3					1	1	30	1	0.1657	0.1899
2			2	2	1			1	30	1	0.1657	0.1899
						1	3	1	27	1	0.1492	0.1709
2		3			2		1	1	27	1	0.1492	0.1709
		3				1	1	1	25	1	0.1381	0.1582
2				3	2			1	23	1	0.1271	0.1456
		2				1	3	1	21	1	0.1160	0.1329
2		3		1				1	19	1	0.1050	0.1203
				2	1			1	18	1	0.0094	0.1139
				3	2		2	1	18	1	0.0094	0.1139

Table 6
Decision rules of negative sales growth ratio with minimum support value ≥ 3 .

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d_1	Support	Certainty	Strength	Coverage
				3	2		2	2	5	1	0.0276	0.2174
2				3			2	2	4	1	0.0221	0.1739
				3			2	2	3	1	0.0166	0.1304
	2			3	2	2		2	3	1	0.0166	0.1304
2	1		2		1	2		2	3	1	0.0166	0.1304
				3		2	2	2	3	1	0.0166	0.1304

element within the complex R&D innovation processes. The important meanings and their substantial contributions to R&D innovation are also identified in the diagram.

Employing the definitions of flow network graphs and the RST introduced in the previous section, we can represent our decision rules by means of flow graphs (Figs. 1 and 2). Figs. 1 and 2 provide presentations of the relationships and paths of the decision flow graph. Flow graphs can be viewed as decision algorithms and each branch can be used to describe a decision rule. According to the decision rules results of RSTs, 125 decision rules that supported a positive sales growth ratio were generated, which indicates the high-technology firms employing these rules in R&D innovation efforts can increase their sales volume. In total, 79 decision rules supporting a negative sales growth ratio were produced, which indicates that the high-technology firms employing these rules in R&D innovation efforts may encounter failures and a lack of sales growth. We can view the entire flow network graph as a decision algorithm, where each branch describes a decision rule. However, in practice, it is too complex to illustrate the relationship among the characteristics of R&D and sales growth if all rules from Tables 5 and 6 are considered. To reduce the complexity of the flow network graph, we selected the top 5 rules (sum

of support equal to 139) and top 2 rules of negative (sum of support equal to 9) sales growth to provide clearer decision-making information.

The coefficient of certainty, strength, and coverage associated with each branch in the flow network graph are illustrated in Table 5. We can see that under the different decision rules, the rule set generates relative strength and coverage. According to the strength and coverage of the decision rules, we can compute and translate all branches into the flow network graph represented in Fig. 1. In Table 5, the coefficient of strength simply represents the ratio of total flow through the branch, while the coverage is used to exhibit how inflow is distributed between inputs of a node.

We have applied decision rules for further translation into the decision algorithms represented in the flow networks graphs illustrated in Fig. 1. Figs. 1 and 2 are the results of the original database of high-technology firms using the RST mining technique to extract series R&D innovation decision rules. Extending the logic of RST and the flow network graph, these states could be viewed as a decision rule in Figs. 1 and 2. Thus, we used the RST decision rule sets in the algorithms for diagnosing and extracting R&D innovation decisions to increase the diagnostic performance and provide useful information for such algorithms. The flow

network graph should therefore incorporate decision rules in algorithms used for the diagnosis of R&D innovation activities. Fig. 1 illustrates the network structure of patterns with the different decision rules. More specifically, this network graph is a decision algorithm connecting R&D innovation decisions with certain conditions. The flow network graph also permits us to identify the dependent factors of R&D innovation and to discover their relationships as a reference for R&D decision makers. We can see from the flow network graph that there is a relatively strong positive dependence of sales growth on R&D innovation decision rules.

Based on the same procedure, we can obtain the flow network graph of decision rules for negative sales growth, as illustrated in Fig. 2. Fig. 2 is based on the top 2 rules from Table 6, in which there is no $a_6=1$. Thus, the decision rules used in the RST were used to reveal that certain classes of decision algorithms can be represented as a flow network graph.

The flow network graph permits for nonlinear relationships with which to analyze the influence of R&D innovation activity on the sales growth of high-technology firms. As mentioned previously, the analyses also provide the possible path and useful information of R&D effort on the degree of dependencies. The dependencies depicted in the flow network graph would be helpful to improve learning on how to accelerate the translation from R&D resource inputs to sales growth for the high-technology firms. Therefore, these flow networks and decision algorithms are valuable for identifying whether there are possible paths and practices that ensure that there exists an appropriate set of decision rules for R&D innovation activities.

5. Conclusions and remarks

We have presented a new approach to predicting R&D innovation efforts in the high-technology industry using RSTs theory and flow networks. The RST was used to analyze and extract hidden decision rules from the initial information. The result of RST is a set of decision rules that can be used to explore undiscovered rules and characteristics. The decision rules can thus be transferred into a flow network graph, used for modeling a flow of information as a set of decision rules and explaining the corresponding flow network in terms of flow distribution. The flow network graph is a bridge for connecting the pathway of decision rules and the degree of their interdependency. It is a useful tool and approach to use for exploring and discovering the path dependences of decision rules, which can permit R&D innovation managers to derive and test predictions about how R&D innovation efforts contribute to sales growth. Most importantly, the RST approach and flow network graph can be used to find patterns in the original data and dependencies between some data structures for subsequent strategic implementation.

Our empirical results illustrating R&D innovation as a predictor of sales growth has illustrated that RSTs are an effective tool for supporting R&D innovation decision-making. The RST approach has revealed that the R&D innovation effort has had a significant impact on sales growth. Our study has found some common patterns, which could permit those managing R&D innovation to allocate R&D resources, select available R&D innovation strategies, and predict potential trends. In light of the empirical implementation, we can see that the RSTs approach is quite robust and simple, particularly in the areas of forecasting and classification decisions, since this approach does not require the pre-specification of a functional form or any particular statistical distribution assumptions about the variables of the model.

Based upon the evolutionary theories of economic and technological change, R&D innovation is an evolutionary and social processes (Edquist, 2004), and the collective learning processes as well as external collaborations with other firms (Cooke et al. 2000). In the same vein, Cooke et al. (1997) argued that the evolution approach is well suited to the analysis of innovation practices because of its emphasis upon process, learning and cooperative, as well as competitive, dimensions of inter-firm relations. Therefore, the clustering of firms in related industries is widely believed and accepted in both academic and policy circles to facilitate innovative activities and promote regional growth, particularly in high-tech industry (Lee, 2009). More specifically, the R&D innovation processes are institutionally embedded in the setting of systems of production. Furthermore, the firm-specific competencies and learning processes can lead to regional competitive advantages in which they are based on localized capabilities such as specialized resources, skills, institutions and share of common knowledge and marketing ability. We focused on the level of competitiveness in the local market. One of the surprising findings is the non-science park firm have significant positive effect on sales growth. This result implies that science park location does not lead to specific advantages in R&D innovation production activities. Additionally, science park high-tech firms did not sufficient deploys their R&D resource allocation, including R&D expenditures, proportion of R&D researchers, and proportion of export in the most appropriate R&D innovation activities. All these constraints impede the high-tech firm's market advantages when the firm R&D- and market-based innovations capability cannot shorten the product life-cycle development may impact on its sales growth.

The advantages of the RST hybrid flow network graph in the R&D innovation are summarized in two points. The first is that the R&D managers and staff can discover hidden information in terms of R&D innovation and predict and act upon the new information based on large-scale initial R&D data. These decision rules can yield a set of dependence paths and rules and can thus lead to a shortening of the R&D processes, high probability of R&D success, and improvement in performance. The second point is that such a model is to be welcomed for its ability to capture the effect of R&D innovation efforts on sales growth and turn this information into a useful innovation device that can be used to check the repeatability of R&D processes and provide a template for R&D innovation benchmarking. This benchmark was offered as a basis and as a useful tool with which high-technology firms can examine and analyze their own R&D processes and practices across their own R&D resources and inputs, eventually gaining a competitive advantage.

The dataset in our study had two limitations. Firstly, we based our findings on a sub-section of the high-technology industry database. As a result, all analyses accounted for high-technology industry characteristics, which may be too narrow to apply to all industries. Although this research defined R&D innovation performance using several R&D-related measures, it is widely recognized that many firms pursue multiple goals and should not only be viewed as profit seekers. As a result of our limitations, we recommend that future research examine certain structural contexts, such as administrative efficiency, geographical propinquity, and domestic-based networks to yield new insights into the impact in the R&D innovation processes. Since a high-technology firm's structural context may play important roles in R&D success, a firm's propensity to repeatedly learn and adjust with certain direction may bring synergy in the firm's overall R&D innovation success. Another limitation of our study is that by using only high-technology firm data, we were constrained by the limitations of considering R&D activity in a single industry level of analysis. For the purpose of accuracy and simplicity, we only studied the high-

technology industry in Taiwanese. Therefore, our findings cannot be generalized to other industries (such as pharmaceutical industry) or countries. As a result, we recommend that future research extend this approach to the different industries to develop a more fine-grained perspective to enhance our understanding of the R&D innovation performance processes. Considering various industries at a more comprehensive view and examining relatedness among them would help capture R&D innovation decision rules more precisely.

This study helped to shed light on rule-based decision-making in mining R&D innovation processes. The empirical findings suggested several implications for research on R&D innovation and causal-and-effect relationships. Overall, the study argued the importance of R&D innovation processes for performance, be it developed by characteristics and components of original data, by the rule-based decision-making technique or from high-technology industry sources. It also illustrated the strong relationship among the various paths and sources of R&D innovation, which underlined the necessity for high-technology firms of having a broad and coherent policy approach for R&D and innovation efforts.

In summary, four implications were drawn in this study. First, a better understanding of the relationship between R&D innovation investment and firms' performance, such as sale growth at the firm level, may further contribute to the design of a more appropriate program, as well as to access sources of R&D innovation activities information. More specifically, high-technology firm enrichment in the R&D information may induce R&D uncertainty and increase innovation success. Second, high-technology firms should provide an appropriate means of R&D performed in the series of R&D innovation activities, in particular, the funding of R&D, R&D researchers, and specific-firm location advantages, which may continue to increase sales growth. Obviously, the R&D innovation investment needs to take into account interdependencies among measurement indicators to ensure that the proper assessment of the R&D innovation is accurate. Third, acquiring firm-specific R&D resources is important for sale growth and competitiveness. High-technology R&D firms should have high absorbing effects, enhancing the ability of the R&D sector to absorb advanced knowledge, techniques, and experiences coming from abroad and/or from the aggregation of science park performed research. Fourth, decisions regarding R&D innovation causal-and-effect relationships, managers can plan how to deploy their R&D innovation strategies to increase their understanding of the complexities of innovation and its management and eventually maximize the innovation returns from R&D innovation efforts.

Acknowledgments

The authors would like to thank the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC 98-2410-H-415-005. The authors would also like to acknowledge the contribution of the two anonymous referees and the Editor-in-Chief, Dr Linton J., his valuable comments and suggestions have contributed significantly to this paper.

References

- Balkin, D.B., Markman, G.D., Gomez-Mejia, L.R., 2000. Is CEO pay in high technology firms related to innovation? *Academy of Management Journal* 43 (6), 1118–1129.
- Baumol, W.J., Wolff, E.N., 1983. Feedback from productivity growth to R&D. *The Scandinavian Journal of Economics* 85 (2), 147–157.
- Beynon, M., Driffield, N., 2005. An illustration of variable precision rough sets model: an analysis of the findings of the UK Monopolies and Mergers Commission. *Computers & Operations Research* 32 (7), 1739–1759.
- Brown, W.B., Gobeli, D., 1992. Observations on the measurement of R&D productivity: a case study. *IEEE Transaction on Engineering Management* 39 (4), 325–331.
- Butchart, R., 1987. A new UK definition of high technology industries. *Economic Review* 400, 83–89.
- Chen, C.J., Huang, C.C., 2004. A multiple criteria evaluation of high-tech industries for the science-based industrial park in Taiwan. *Information & Management* 41 (7), 839–851.
- Coccia, M., 2001. A basic model for evaluating R&D performance: theory and application in Italy. *R&D Management* 31 (4), 453–464.
- Coombs, J.E., Bierly, P.E., 2006. Measuring technological capability and performance. *R&D Management* 36 (4), 412–438.
- Cooke, P., Uranga, M.G., Etxebarria, G., 1997. Regional innovation systems: institutional and organizational dimensions. *Research Policy* 26, 475–491.
- Cooke, P., Boekholt, P., Tödtling, F., 2000. In: *The Governance of Innovation in Europe*. Pinter, London.
- Chang, P.L., Hsu, C.W., 1998. The development strategies for Taiwan's semiconductor industry. *IEEE Transactions on Engineering Management* 45 (4), 349–356.
- Del Monte, A., Papagni, E., 2003. R&D and the growth of firms: empirical analysis of a panel of Italy firms. *Research Policy* 32 (1), 1003–1014.
- Diaz, M., Gomez-Mejia, L., 1997. The effectiveness of organization-wide compensation in technology-intensive firms. *Journal of High Technology Management Research* 8 (2), 301–315.
- Dimitras, A., Slowinski, R., Susmaga, R., Zopounidis, C., 1999. Business failure prediction using rough sets. *European Journal of Operational Research* 114 (1), 263–280.
- Dosi, G., 1988. Sources, procedures and microeconomic effects of innovation. *Journal of Economic Literature* 25 (1), 1120–1171.
- Dressler, R., Wood, R.S., Alvarez, Z., 1999. Evaluating R&D performance using the cost saving metric. *Research-Technology Measurement* 42 (2), 4–13.
- Duysters, G., Hagedoorn, J., 2000. Core competencies and company performance in the world-wide computer industry. *Journal of High Technology Management Research* 11 (1), 75–91.
- Duysters, G., Hagedoorn, J., 2001. Do company strategies and structures converge in global markets? Evidence from the company industry. *Journal of International Business Studies* 32 (1), 347–356.
- Ettlie, J.E., 1998. R&D and global manufacturing performance. *Management Science* 44 (1), 1–11.
- Edquist, C., 2004. Systems of innovation—a critical review of the state of the art. In: Fagerberg, J., Mowery, D., Nelson, R. (Eds.), *Handbook of Innovation*. Oxford University Press, Oxford.
- Fan, T.F., Liu, D.R., Tzeng, G.H., 2007. Rough set-based logistic for multi-criteria decision analysis. *European Journal of Operational Research* 182 (1), 340–355.
- Ford, L.R., Fulkerson, D.R., 1962. In: *Flows in Networks*. Princeton University Press, Princeton, NJ.
- Franko, L.G., 1989. Global corporate competition: who's winning, who's losing, and the R&D factor as one reason why. *Strategic Management Journal* 10 (1), 449–474.
- Graves, S.B., Langowitz, N.S., 1996. R&D productivity: a global multi-industry comparison. *Technology Forecasting and Social Change* 53 (2), 125–137.
- Griffith, R., Redding, S., Van Reenen, J., 2004. Mapping the two faces of R&D: productivity growth in a panel of OECD Industries. *The Review of Economics and Statistics* 86 (4), 883–895.
- Grupp, H., Maital, S., 2000. Perceived innovation of Israel's largest firms: an empirical study. *Technovation* 20 (3), 129–137.
- Guellec, D., Van Pottelsberghe de la Potterie, B., 2004. From R&D to productivity growth: do the institutional settings and the source of funds of R&D matter? *Oxford Bulletin of Economics and Statistics* 66 (33), 353–378.
- Hagedoorn, J., Cloodt, M., 2003. Measuring innovative performance: is there an advantage in using multiple indicators? *Research Policy* 32 (1), 1365–1379.
- Hall, B., Mairesse, J., 1995. Exploring the relationship between R&D and productivity in French manufacturing firms. *Journal of Econometrics* 65 (1), 263–293.
- Hartmann, G.C., Mark, B.M., Richard, S.R., 2006. Planning your firm's R&D investment. *Research Technology Management* 49 (1), 25–36.
- Henderson, R., Cockburn, I., 1994. Measuring competence: exploring firm effects in pharmaceutical research. *Strategic Management Journal* 15 (1), 63–84.
- Ito, K., Pucik, V., 1993. R&D spending, domestic competition, and export performance of Japanese manufacturing firms. *Strategic Management Journal* 14 (1), 61–75.
- Keeble, D., Wilkinson, F., 2000. In: *High-technology clusters, networking and collective learning in Europe*. Ashgate, Aldershot, VT.
- Kim, B., Oh, H., 2002. An effective R&D performance measurement system: survey of Korean R&D researchers. *Omega-International Journal of Management Science* 30 (1), 19–31.
- Koschatzky, K., Bross, U., Stanovnik, P., 2001. Development and innovation potential in the Slovene manufacturing industry: analysis of an industrial innovation survey. *Technovation* 21 (5), 311–324.
- Lee, J., Shin, E., 1995. Moderating effects of R&D on corporate growth in U.S. and Japanese hi-tech industries: an empirical study. *The Journal of High Technology Management Research* 6 (2), 179–191.

- Lee, C.Y., 2009. Do firms in clusters invest in R&D more intensively? Theory and evidence from multi-country data. *Research Policy* 38, 1159–1171.
- Long, W.F., Ravenscraft, D.J., 1993. LBO's debt and R&D intensity. *Strategic Management Journal* 14 (1), 119–135.
- Mark, B., Munakata, T., 2002. Computing, artificial intelligence and information technology. *European Journal of Operational Research* 136 (1), 212–229.
- Medcof, J.W., 1999. Identifying 'super-technology' industries. *Research Technology Management* 42 (1), 31–36.
- Morbey, G.K., Reithner, R.M., 1990. How R&D affects sales growth, productivity and profitability. *Research Technology Management* 33 (3), 11–14.
- OECD, 1993. In: *The Measurement of Scientific and Technological Activities: Standard Practice for Surveys of Research and Experimental Development – Frascati Manual 1993*. OECD, Paris.
- Ou Yang, Y.P., Shieh, H.M., Tzeng, G.H., Yen, L., Chan, C.C., 2008. Business aviation decision-making using rough sets. In: Chan, C. –C. (Ed.), *RSTC 2008*, 5306. LNAI, pp. 329–338.
- Pawlak, Z., 1982. Rough set. *International Journal of Computer and Information Sciences* 11 (1), 341–356.
- Pawlak, Z., 1991. *Rough sets, Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht.
- Pawlak, Z., Grzymala-Busse, J., Slowinski, R., Ziarko, W., 1995. Rough sets. *Communications of the ACM* 38 (11), 89–95.
- Pawlak, Z., 1996. Why rough sets? *The Fifth IEEE International Conference on Fuzzy Systems* 2, 738–743.
- Pawlak, Z., 1997. Rough sets. In: Lin, T. y., Cercone, N. (Eds.), *Rough Sets and Data Mining-Analysis or Imprecise Data*. Kluwer, Boston, MA.
- Pawlak, Z., 2002. Rough sets, decision algorithms and Bayes' theory. *European Journal of Operational Research* 136 (1), 181–189.
- Pawlak, Z., 2004. Decision rules and flow networks. *European Journal of Operation research* 154 (1), 184–190.
- Pawlak, Z., 2005. Flow graphs and intelligent data analysis. *Fundamenta Informaticae* 64 (1), 369–377.
- Porter, M.E., 1980. In: *Competitive Strategy*. Free Press, New York.
- Porter, M.E., 1990. In: *The Competitive Advantage of Nations*. Free Press, New York.
- Qian, G., Li, L., 2003. Profitability of small- and medium-sized enterprises in high-tech industries: the case of the biotechnology industry. *Strategic Management Journal* 24 (9), 881–887.
- Reeble, D., 1990. High technology industry. *Geography* 75, 361–364.
- Roper, S., Love, J., 2002. Innovation and export performance: evidence from the UK and German manufacturing plants. *Research Policy* 31 (7), 1087–1102.
- Shyng, J.Y., Wang, F.K., Tzeng, G.H., Wu, K.S., 2007. Rough set theory in analyzing the attributes of combination values for the insurance market. *Expert Systems with Applications* 32 (1), 56–64.
- Smith, V., Dilling-Hansen, M., Eriksson, T., Madsen, E.S., 2004. R&D and productivity in Danish firms: some empirical evidence. *Applied Economics* 36 (1), 1797–1806.
- Sterlacchini, A., 1999. Do innovative activities matter to small firms in non-R&D-intensive industries? An application to export performance. *Research Policy* 28 (8), 819–832.
- Tay, F.E.H., Shen, L., 2002. Economic and financial prediction using rough sets model. *European Journal of Operation Research* 141 (1), 641–659.
- Thornhill, S., 2006. Knowledge, innovation and firm performance in high- and low-technology regimes. *Journal of Business Venturing* 21 (1), 687–703.
- Tsai, K.H., Wang, J.C., 2004. The R&D performance in Taiwan's electronics industry: a longitudinal examination. *R&D Management* 34 (2), 179–189.
- Wakelin, K., 2001. Productivity growth and R&D expenditure in UK manufacturing firms. *Research Policy* 30 (1), 1079–1090.
- Walwyn, D., 2007. Finland and the mobile phone industry: a case study of the return on investment from government-funded research and development. *Technovation* 27 (1), 335–341.
- Wan, D., Ong, C.H., Lee, F., 2005. Determinants of firm innovation in Singapore. *Technovation* 25 (3), 261–268.
- Wang, C.H., Lu, I.Y., Chen, C.B., 2008. Evaluating firm technological innovation capability under uncertainty. *Technovation* 28 (1), 349–363.
- Wang, J.C., Tsai, K.H., 2003. *Productivity Growth and R&D Expenditure in Taiwans' Manufacturing Firms*. NBER Working Paper. National Bureau of Economic Research Inc.
- Werner, B.M., Souder, W.E., 1997. Measuring R&D performance—state of the art. *Research-Technology Measurement* 40 (2), 34–42.
- Willigers, B.J.A., Hansen, T.L., 2008. Project valuation in the pharmaceutical industry: a comparison of least-squares Monte Carlo real option valuation and conventional approaches. *R&D Management* 38 (5), 520–537.
- Wu, M.C., Yen, S.H., 2007. Pricing real growth options when the underlying assets have jump diffusion processes: the case of R&D investments. *R&D Management* 37 (3), 269–276.