國立交通大學

科技管理研究所

博士論文

生活型態在實體與虛擬通路之市場區隔研究

An improved novel model based on lifestyle perspective for market segmentation of physical and virtual channels

研究生：陳瑾儀
指導教授：曾國雄 講座教授

中華民國九十八年六月十六日

生活型態在實體與虛擬通路之市場區隔研究
An improved novel model based on lifestyle perspective for market segmentation of physical and virtual channels

研 究 生：陳瑾儀　　Student：Chin-Yi Chen

指導教授：曾國雄　　Advisor：Gwo-Hshiung Tzeng

國 立 交 通 大 學

科 技 管 理 研 究 所

博 士 論 文

A Dissertation

Submitted to Department of Management

Graduate Institute of Management of Technology

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Management

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

生活型態在實體與虛擬通路之市場區隔研究

學生：陳瑾儀　　指導教授：曾國雄

國立交通大學科技管理研究所 博士班

摘　　要

在動態的環境下，因為可以接觸到的商品種類與傳播媒體的發達，造成消費者的偏好分歧，也使得以往傳統的市場區隔愈加難以解釋消費者選擇行為。本研究希望改善以往的市場區隔模式，首先利用生活型態變數來增加解釋的豐富性，並提出一個結合一般生活型態變數與適當的方法論的新市場區隔模型來解決之前模式的複雜性。在實證分析中，為了簡化過去生活型態問卷的長度與繁雜度，本研究分別針對實體與網路購物行為，利用分類與決策樹與約略集合理論來找出與傳統研究不同的市場區隔變數。本研究發現，利用生活型態變數所找出的市場區隔之預測購買行為之能力，除了可以找出不同於以往的市場區隔之內涵外，分類與預測之效果可比傳統人口統計變數佳；此外，分類與決策樹加上約略集合理論兩種研究方法的結合，將可以有效加強此模式分類與預測之結果。

關鍵字：一般生活型態變數，分類與決策樹，約略集合

An improved novel model based on lifestyle perspective for market segmentation of physical and virtual channels

student：Chin-Yi Chen　　Advisors：Dr. Gwo-Hshiung Tzeng

Graduate Institute of Management of Technology
National Chiao Tung University

ABSTRACT

Under the dynamic environment, consumers have much more chance to access products and media; this may create more difference in their preferences. This study aims to improve the elder method of market segmentation; first, lifestyle variables are used to predict consumer choices to increase the richness of the explanation. Second, the novel model which integrates general lifestyle variables and appropriate methodologies is proposed to solve the complexity of the previous research problems. To simplify and reduce the length of the questionnaire in lifestyle research, classification and regression tree (CART) is applied to discover the general lifestyle variables to segment the market. The empirical results show that, under different purchasing situation, the explanatory power of general lifestyle variables is not less than those traditional demographic variables; in fact, some models with different combinations of general lifestyle variables have higher explanatory power than demographic ones. Also, in this paper, a novel method, integrated by CART and rough sets, is propose to improve the shortcoming of CART. From the comparison of CART, rough sets and the proposed method, we can conclude that the proposed method appropriately integrates the advantages of CART and rough sets.

Keywords: General lifestyle variables, CART, Rough set

誌　　謝

在漫長的博士班就讀期間，有很多很多需要感謝的人。

在完成論文的過程中，首先要感謝的是我的指導教授曾國雄老師在就讀博士班期間耐心的照顧與付出，一直不斷的鼓勵我讓我有信心把學業完成；鄭老師(CEO)給我的研究方向，與五年來的信任與協助，讓我能順利的完成學業。以及科管所的徐老師、洪老師、林老師、袁老師、虞老師，非常謝謝你們在各方面的照顧。JJ 學長謝謝你在論文上的幫忙與教導，也不斷的督促我把學業一股作氣的完成，真的很謝謝你。除此之外，我要感謝家人的支持與付出，以及這五年來的包容，如果沒有你們，我是不可能會有今天的。

在此刻想要特別感謝最美麗的榆淨與曉琪學姊，謝謝妳們這幾年來無私的幫助與陪伴，也謝謝你們帶給最珍貴的友情，在我最無助的時候幫助我一步步的度過難關，沒有妳們就不會有現在的我。此外，還有雅雯學姊、A 王學長、小王子宗偉學長、胡宜中學長、朱克聰學長，謝謝你們這一路走來的幫助與鼓勵，與在科管所的同學又心、麗敏、司令你們的陪伴也讓在科管所的生活更加的順利與開心，還有讓我留下許多難忘的回憶的 Frederic。也謝謝中原企管系的王主任、雅惠、小賴、秋月姊在中原的照顧，在忙的時候幫助我給我很多空間能夠專心的把論文寫完。

很感謝 G5 的牛哥、王大哥對我在工作上的提攜與照顧、以及可愛的 SOAR 四千金 Joy、Ginny、Mickey、Candy，跟你們一起討論、談心的時光真的很棒，我們的友誼一定要繼續下去喔~還有很有趣的雪芳、弘毅，謝謝你們一路走來的陪伴，因為有你們，才讓我有一直堅持下去的勇氣。

最後，還有一直陪在我的身邊，陪伴我走過最艱困的時光小花，相伴四年的小黑，熱心又可愛的 wen，以及我的高中死黨明禪、仙、柚子，還有旺旺、雅茹、勳爺、林董、威董、佳山、凱哥、至勳、Benson、小帥、彥廷、Redrain、RT 學長、lulala、律芸、宸釗、GoGo，很開心這一段求學的人生中有你們的陪伴，讓我在這五年的日子裡一點也不孤單，謝謝你們，也希望接下來的日子，大家都能夠一切順利、平安。

Table of content

Table captions

Figures captions

Chapter 1 Introduction

1.1 Research Background and motivation

With the growing complexity of the society, consumer preference has become more unpredictable because of more sources of information collection and more products choices. To enhance the capability of competition on the market, to know customer more has always the most important mission to companies. In marketing research, the explanatory value of traditional criteria is in steady decline. This is because the individuals who together make up the market every day provide examples of very similar purchase behavior patterns on the part of people differing considerably in social-economic and demographic terms and vice versa, with growing personalization of consumer habits being observed (Gonzalez and Laurentino, 2002). Therefore, the deep and wide-ranging changes which present-day society is undergoing must also be taken into account.

This study will separate into two parts according to the different channels that consumer can shop through them, that is, virtual channel and physical channel. The former is the channel that consumer should shop by going to the “brick and mortar” store, the latter is about internet online shopping. There are many lifestyle researches have analyzed consumer behavior of different product targets in the physical channel, for instance, food, red wine or trip choices. However, the lifestyle approach those researches used mostly are product-specification, which means the usage of the research is limited, i.e., the results of the analysis will be hard to apply to other product categories later. And also, their model of doing the marketing research has its

own drawbacks. For instance, questionnaire is often too long to make sure respondents provide the right answers or the research method could miss or neglect huge amount of information form the data which is collected.

To solve the problems, this research aims to improve the traditional research model, through literature review and expert advice. A general lifestyle structure is going to be generated, after the appropriate methodologies are applied, the patterns of consumer behavior will be discovered, and that is, the decent linkages between consumer's lifestyle and their product choice can be found. After obtaining the result of analysis, we can deduce the marketing strategies and suggestions to the companies.

Instead of physical channel, virtual channel such as internet has become an important way of shopping nowadays. Internet provides new opportunities for sellers and buyers. With the increasing number of people shop thought this channel, the market have enjoyed a rapid growth. Business-to-consumer electronic commerce is growing in every category of goods, for instance, financial services, online-travel services, computer hardware and software, book and music all accompanied with good sales performance in these years (Kim et al., 2001). However, due to the characteristics of the e-commerce industry, companies with limited resources face extremely high competition, to discover effective means of marketing online will always be a critical issue to internet stores. To achieve the maximum efficiency of marketing, the key to success will rely on the market segmentation, that is, the important fundamental marketing analytic basis first proposed by Wendell (1956). Since companies need to satisfy the diverging preferences of customers and react to the complex online business environment quickly, the new marketing strategies such as one-to-one marketing have been stressed by researches and practical affairs. One to one marketing (also known as database marketing) introduces a fundamental new

basis for competition in the marketplace by enabling organizations to differentiate based on customers rather than products (Peppers and Rogers, 1993). The main purpose of one to one marketing is not market share but finding a group of valuable customers toward specific products. One solution to realize these strategies is personalized recommendation that helps customers find the product according to their interests by producing a list of products for each given customer (Cho et al., 2002). That is, an effective way to increase customer satisfaction and consequently customer loyalty.

Nowadays, a variety of recommendation techniques has been developed and recommendation systems are commonly applied by B to C companies, for instance, Amazon.com and CDNow.com, they use an intelligent engine to mine the customers' ratings records and then create predictive user models for product recommendation. Typically there are two kinds of recommender systems: content-based and collaborative systems. The former provide recommendations to a customer by automatically matching customers' interests with product contents; the latter, has been known to be the most successful, provide recommendations by utilizing overlap of preference ratings to combine the opinions of "like-minded" customers, that is, identify customers whose interests are similar to those of a given customer have liked.

Until now, there are researches about one to one marketing, that is, using data mining techniques to segment the market and forming the recommendation systems use demographics variable as the only basis of market segmentation, that is, lifestyle and values are not involved in the research. And also, many lifestyle researches about online shopping; however, most of their target is to find a wired-lifestyle, that is, using life-related variables to analyze if people shop online or not. Concerning the description above, to add lifestyle and values as variables to segment the market could

enrich the recommendation systems. Therefore, to increase the accuracy of matching the customer and products online, the general lifestyle structure is worth to be included. Since the variables used to segment the market of recommend system are not only demographics, the understanding of customers can be improved and the efficiency of the interaction with them will be increased hopefully.

In this study, questionnaire design requires special data mining tools which and deal with different data scales, considering several methodologies, rough set theory (RST) and classification and regression tree (CART) are used in this study to analyze the content and features of data.

Classification and regression tree, developed by (Breiman et al., 1984), is a flexible and robust analytical method, which can deal with nonlinear relationships, high-order interactions, missing values, and this method is simple to understand and give easily interpretable results. There are several advantages of CART: (1) the flexibility to handle a broad range of response types, including numeric, categorical, ratings, and survival data; (2) invariance to monotonic transformations of the explanatory variables; (3) ease and robustness of construction; (4) ease of interpretation; and (5) the ability to handle missing values in both response and explanatory variables. Thus, trees complement or represent an alternative to many traditional statistical techniques, including multiple regression, analysis of variance, logistic regression, log-linear models, linear discriminant analysis, and survival models. Rough set theory, which was developed by Pawlak (1982), is a rule-based decision-making technique that can handle crisp datasets and fuzzy datasets without need for a pre-assumption membership function. It can also deal with uncertain, vague, and imperceptible data.

The remainder of this study is organized as follows. Section 2 describes the literature review to be the fundamental of the study. Section 3 describes the methodology of classification and regression tree and rough set theory. In section 4,

two real cases such as the purchasing behavior of digital camera and books online are presented to show the new segmentation process by CART and RST. Finally, in section 5, we present the conclusions and suggestions.

1.2 Research Objectives

Under the dynamic environment, consumers have much more chance to access products and media; this may create more difference in their preferences. Therefore, no matter marketing in physical or in virtual channel, to make effort to achieve higher accuracy of market segmentation is necessary. The traditional analytic model which based on factor analysis should be improved to match the needs from the dynamic market scenarios

This study aims to improve the elder method of market segmentation. As addressed above, lifestyle and values are important variables to influence consumer choices that should be included in the research to increase the richness of the explanation. In virtual channel, moreover, the model we presented here could be a new basis to segment the market online. Since most of the study of lifestyle has its own AIO measurement, this study also aims to create the survey with rich information and appropriate length. That is, using data mining tools to find the variables that have the critical influence to consumer choices. Therefore, to cut down the length of questionnaire and develop one moderate general lifestyle structure specific for Taiwan is also an important objective in this research.

After the conceptual structure of the basis in this study, market segmentation for different product targets are going to be discovered and explained. The link between consumers' personal differences and product choice will be established, and their preferences in different segments are going to be discussed. After having more information of consumers in this research construct, decent market strategies in different channels can be generated. The comparison will be made to check if lifestyle and values variables can increase the accuracy of market segmentation in each market.

Since there are two rule-based data mining techniques applied here, the result of comparison will be concluded at the end.

To provide different market segmentation model, as the description mentioned above, there are three main objectives in this research:

(1) To develop a moderate survey of lifestyle and values for Taiwan;

(2) To establish the effective linkage between consumers lifestyle and product choices;

(3) Proposing this novel model to segment the market in different channels.

Chapter 2 Literature Review

In this chapter, main concepts of this research from previous literature will be presented, includes consumer behavior, market segmentation and the construct of lifestyle, that is, activities, interests and opinions. Also, other similar research used nowadays will also be proposed in this chapter.

2.1 Consumer behavior model

2.1.1 Definition of consumer behavior

Consumer behavior is a theory that integrates many different theories from different research field, for instance, marketing, economics, psychology and social science. Therefore, many scholars had proposed definitions of it.

From Engel et al. (1995): consumer behavior is those actions directly involved in obtaining, consuming, and disposing of products and services, including the decision process that precede and follow these actions.

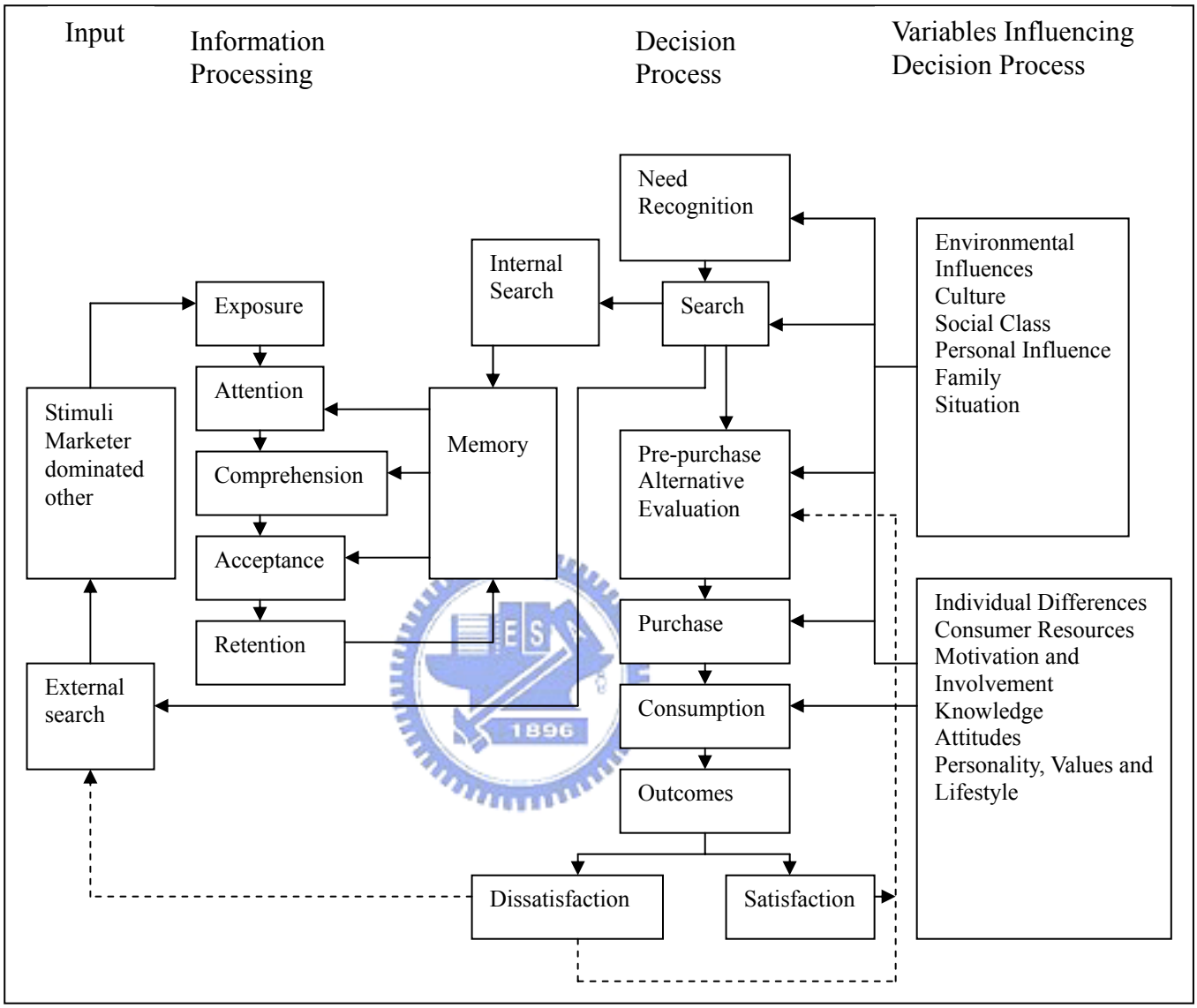
In this section, the consumer behavior model from Engel-Kollat-Blackwell is used as the basis of the concept in this research; this model is described as follows.

2.1.2 Models for consumer behavior: Engel-Kollat-Blackwell Model

Engel-Kollat-Blackwell model are proposed in 1968, after being revised in seven times, has became a complete and systematic structure in consumer behavior theory.

This model confers consumer behavior from the decision process. Integrated with the opinions from elderly scholars, it has become a full-scale model which regards consumer behavior as a sequential process instead of discontinuous individual actions. The characteristic of this model is it pus decision process as its central concept and also combines inner and outer factors interchangeably to create a whole decision making system.

There are fore main components in EKB model: input, information processing, decision process and variables influencing decision process.



Source: Engle, James F., Blackwell, Roger D. & Miniard, Paul W., Consumer Behavior, 8th ed. Orlando Florida, Dryden Press, P.263, 1995.

Fig. 1 Engel-Kollat-Blackwell model

Kolter (1998) believes that the research of consumer behavior is to understand the whole process of black box. To know more about this implicit course, we have to recognize the process of consumer's decision making and the background and characteristics of him. Besides, EKB model has detailed discussion about the process of decision making and the sources of the factors which influence his decision making. From this model, it is obvious that individual difference plays an important role to the decision making process, that is, it influences consumer choices directly or indirectly, since the main purpose of this study is to

find out how personality and lifestyle, these individual differences, can affect consumers choices, therefore, this model is the appropriate conceptual base of this study.

2.2 Market segmentation

2.2.1 Segmentation to Marketing strategy

In today's market, there are fewer and fewer situations where a mass marketing approach is feasible. This has come as a direct result of the increasing diversity of consumer needs (Wedel and Kamakura, 2000). The justification for segmenting consumer markets is that consumers who share similar characteristics will share similar attitudes, wants and needs and therefore responses towards marketing stimuli (Ahmad, 2003; Dibb et al., 2002). Whereas heterogeneity is probably the most important reason for segmentation (Hunt and Arnett, 2004), decades age Smith (1956:5) defined market segmentation as "viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desires of consumers for more precise satisfaction of their varying wants". Segmentation can be based on situations, product-situation and person-situation interactions (Van Raaij and Verhallen, 1994).

The nexus of market segmentation is that it allows a business to deal with diverse customer needs in a resource-efficient manner (Dibb and Simkin, 1996). The utility of market segmentation is hence rooted in two main reasons; namely, it enables the need analysis of a specific consumer segment, and it fosters marketing campaigns to be focused on the identified needs (Thach and Olsen, 2006).

Since segmentation have been the most frequent topic covered under the generic heading of marketing management, the role of segmentation in marketing strategy is presented in here.

For segmenting consumer and business markets, the major segmentation variables –geographic, demographic, psychographic, and behavioral segmentation- are summarized in Table 1.

Table 1 Major segmentation variable for consumer markets

Segmentation types	Segmentation variables
Geographic	Region; City or metro size; Density; Climate

Demographic	Age; Family size; Family life cycle; Gender; Income; Occupation; Education; Religion; Race; Generation; Nationality; Social class
Psychographic	Lifestyle; Personality
Behavioral	Occasions; Benefits; User rate; Usage rate; Loyalty status; Readiness stage; Attitude toward product

Source: Kolter, P., 2000. Marketing Management. pp. 264.

Segmentation is a critical work for company before planning marketing strategy. Therefore, adequate market-segmentation procedure is useful. There is three-step procedure for identifying market segments (Kolter, 1997).

Step 1: Survey stage

The researcher conducts exploratory interviews and focus groups to gain insight into consumer motivations, attitudes, and behavior. Then the researcher prepares a questionnaire and collects data on attributes and their important information.

Step 2: Analysis stage

The researcher applies factor analysis to the data to remove highly correlated variables, and then applies cluster analysis to create a specified number of maximally different segments.

Step 3: Profiling stage

Each cluster is profiled in terms of its distinguishing attitudes, behavior, demographics, psychographics, and media patterns. Each segment is given a name based on its dominant characteristic.

The process of market segmentation in this study will follow these stages presented above, and demographics, lifestyle and value these variables are used to segment the target market, with appropriate methodologies, we believe that the accuracy of marketing strategies generated for different market segments can be improved.

2.3 Lifestyles and Values

2.3.1Definition of psychographics and lifestyles

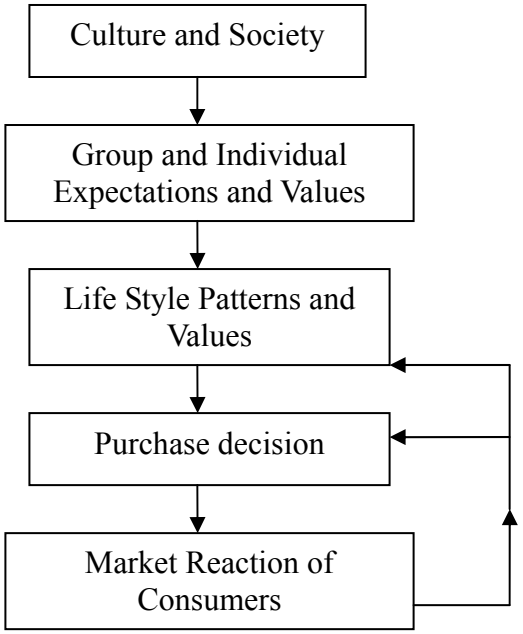
Lifestyle variables are usually associated with psychographic segmentation methods (Heath, 1995; Grunert et al., 1997), meaning that the variable… "was subjective in nature (in contrast to objective criteria like demographics or rate of usage) and that it was not product specific (unlike product attitudes or preferences), but rather a general characteristic of the consumer" (Grunert et al., 1997:4, Askegaard & Brunsø, 1999). Until now, the terms "lifestyles" and "psychographics" are used interchangeably in the marketing literature, and there is much overlap in what these terms are generally thought to mean. One distinction between two terms is that the main purpose of psychographics is to obtain a better understanding of the consumer as a person by measuring him on multiple psychological dimensions, for instances, Nelson, Dorney and Peterson have used the term "psychographics" to refer to studies that place comparatively heavy emphasis on generalized personality traits.

The core concept of life style mainly derived from psychology and social science. Since 1960s, psychographics and lifestyles have received wide attention among consumer researchers. The application of lifestyle variables in consumer research had first been introduced by Lazar in 1963 (cited by Grunert et al., 1997; Kamakura & Wedel, 1995; Wedel & Kamakura, 2000). From Lazar, the definition of life style is as follows:

Life style is a systems concept. It refers to the distinctive or characteristic mode of living, in its aggregative and broadest sense, of a whole society or segment thereof. It is concerned with those unique ingredients or qualities which describe the style of life of some culture or group, and distinguish it from others. It embodies the patterns that develop and emerge from the dynamics of living in a society.

Lifestyle, therefore, is the result of such forces as culture, values, resources, symbols, license, and sanction. From one perspective, the aggregate of consumer purchases, and the manner in which they are consumed, reflect a society's life style.

In conjunction with the definition, Lazar offers the "life style hierarchy" diagrammed below (see Fig. 2):



Source: Thomas P. Hustad & Edgar A. Pessemier, "The Development and Application of Psychographics", in William D. Wells ed., Life Style and Psychographics, Chicago AMA, p.37, 1974.

Fig. 2 Lazar life style hierarchy

2.3.2 AIO statements

Early lifestyle segmentation studies operationalized the life-style construct through a large battery of Likert-type statements covering the following categories (Plummer, 1974).

Table 2 Life style dimensions

Activities	interests	Opinions	Demographics
Work	Family	Themselves	Age
Hobbies	Home	Social issues	Education
Social events	Job	Politics	Income
Vacation	Community	Business	Occupation
Entertainment	Recreation	Economics	Family size
Club membership	Fashion	Education	Dwelling
Community	Food	Products	Geography
Shopping	Media	Future	City size
Sports	Achievements	Culture	Stage in life cycle

Source: Joseph T. Plummer (1974), "The Concept and Application of Life Style Segmentation," Journal of Marketing, Vol.62, May, pp.34-36.

- Activities: Reported behavior related to club membership, community, entertainment hobbies, shopping, social events, sports, vacation and work.
- Interests: Degree of excitement about and attention to achievement, community, family, fashion, food, home, job, media and recreation.
- Opinions: Beliefs about business, culture, economy, education, future, politics, products, self and social issues.

AIO components are also defined by Reynolds and Darden as follows:

An activity is a manifest action such as viewing a medium, shopping in a store, or telling a neighbor about a new service. Although these acts are usually observable, the reasons for the actions are seldom subject to direct measurement. An interest in some object, event, or topic is the degree of excitement that accompanies both special and continuing attention to it. An opinion is a spoken or written "answer" that a person gives in response to stimulus situations in which some question is raised. It is used to describe interpretations, expectations, and evaluations-such as beliefs about the intentions of other people, anticipations concerning future events, and appraisals of the rewarding or punishing consequences of alternative courses of action.

Aside from the Likert-type items about activities, interests and opinions, most AIO researches include demographic variables. The broad range of areas listed above is commonly used in AIO studies that can be applied to more than one product market. Such studies may include between 200 to 300 AIO statements. Commonly, a data reduction technique such as factor analysis is first used to translate the large battery of items into a small, more meaningful and interpretable number of underlying psychographic dimensions (Alpert and Gatty, 1969; Darden and Reynolds, 1971; Reynolds and Darden, 1972; Moschis, 1976). For example, in an early study, Wells and Tigert (1971) used 300 Likert-type items that were then factor analyzed and reduced to 22 lifestyle dimensions. However, there are problems of this approach that researcher started to argue, from Gonzalez and Laurentino (2002): …

Surveys are over-long not just because of the large amount of data but also, perhaps even more, because of the scales used, mostly based on scoring, for each and every item in the questionnaire. This means that the respondents have to stop and think hard so as to rate each single item, leading to great fatigue, often not really necessary. Questionnaires are customarily self-administrated, which is an added handicap, as the questions used are

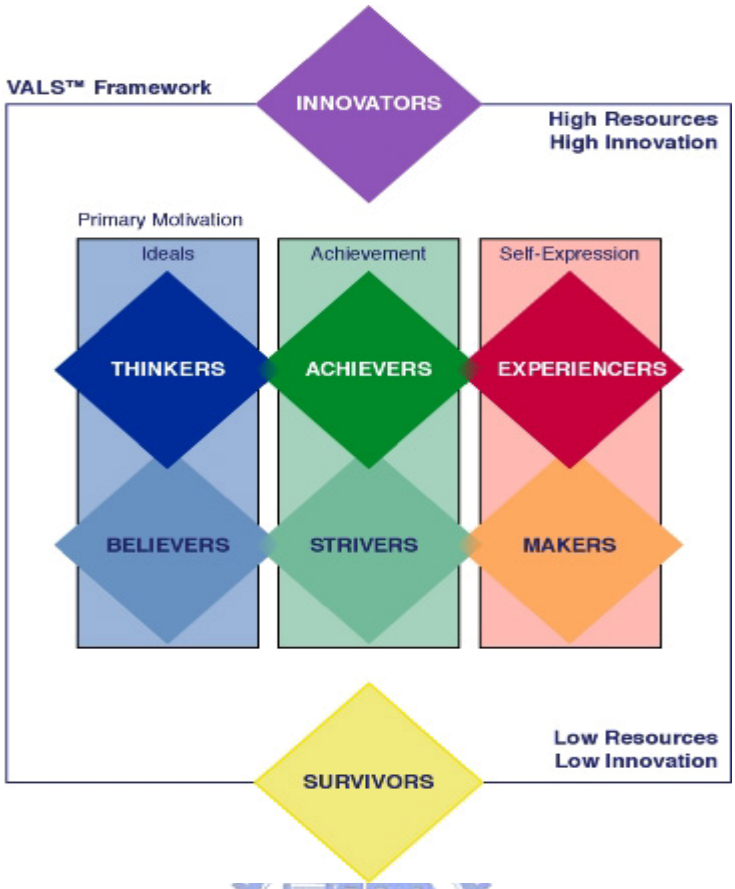
sometimes not easy to understand, and this form of survey does not allow for clarifications. Both these factors, length and difficulty in understanding questions, lead to the spoiling of a great many questionnaires, especially those filled in by elderly folk, and to considerable biasing of data by flagging attention during the long time-span required for completion (Valette-Florence, 1994).

The bulkiness of the ensuing data also triggers a need for multiple statistical analyses, including factorial analysis of principal components, used with the aim of reducing the number of variables. These retain only a part of the information, to which further techniques are applied, thus leading in the end to a considerable amount of unused residual information, as the raw data are not handled directly, but undergo successive impoverishment.

The comments concerning the criticisms brought forward against the variables traditionally used in segmenting the market, particularly their limited explanatory value in the context of developed countries (Ritchie and Goeldner, 1987; Fisher, 1990; Mitchman, 1991; Witt and Moutinho, 1994; Lambin, 1995), justify the aims of the present study. Instead of following the elderly approach, this study are going to be an improvement in the techniques for measuring lifestyle, followed by use of the construct to divide up consumers, provided that a relationship can be demonstrated between lifestyle and consumer behavior.

2.3.3 VALS and LOV

A widely used approach to lifestyle marketing is the values and lifestyle (VALS) and its most recent from, VALS Ⅱ. The original program was developed by Mitchell at SRI and defined nice American lifestyles shown in Fig.3, along with typical demographics and buying patterns. The main dimensions of the segmentation framework are primary motivation (the horizontal dimension) and resources (the vertical dimension). SRI describes consumer market segments as ideals-driven, achievement-directed, and self-expression. The system defines a typology of three basic categories of consumer values and lifestyles, with eight more-detailed types, such as innovators, thinkers, achievers, experiencers, believers, strivers, makers and survivors.



Sources: http://www.sric-bi.com/VALS/types.shtml

Fig. 3 VALS Ⅱ lifestyle segmentation

VALS uses the results of survey to segment the market and discover the suitable type of consumers to match different marketing proposes, until now, has gained rapid acceptance and widespread usage in marketing. However, it has its limitations, for instance, the result can only represent a broad picture of lifestyle temporarily, and also, huge diversity exists in different countries, societies, therefore it's hard to say if this complete measurement of lifestyle is suitable to Taiwanese people, however, since lifestyle doesn't change very often by times and has stabilities, VALS still has the value to be referenced. Some discriminating questions from this approach are going to be collected in this study in the questionnaire through expert advice.

In VALS survey, respondents are given a score that reflects the degree to which they share similar responses on lifestyles other than their primary lifestyle, therefore, consumers are not "pure" in their type of lifestyle. Since VALS is a proprietary data base, some consumer researchers also criticize the act that researchers do not have full information on the factor loadings or rotations or the explained variance, causing them to revert to the more

basic approaches based on the academic value studies of Rokeach and Schwartz.

Another alternative to VALS is the list of value (LOV) approach, developed by Kahle (1983). Typically, respondents are asked to rank a list of values derived from the Rokeach's value survey (RVS), then marketers can use the top-ranked value to assign consumers to segments. The main distinction between the RVS and LOV sacles, from a theoretical point of view, is that the latter dose not include any value related to societal interests.(market segmentation). An empirical comparison between the RVS and LOV scales on a convenience sample of 356 residents of a college town (Beatty et al., 1985) showed limited evidence of convergent validity between the two scales. Also, there are several studies aims to compare VALS and LOV approaches, Kahle et al. (1986) found the LOV approach predicted consumer behavior better than VALS. But from the research of Thomas et al. (1990), when used alone, VALS appears better than LOV, but when demographic data are included with LOV, the latter approach is more effective. Research by Kamakura and Novak (1992) incorporates the more conceptual approach of Schwartz to define market segments on the basis of the latent value systems of market segments, this extension of LOV approach reflects the multiple values that affect an individual's buying behavior, provides a richer understanding of the activities and interests of the segments.

After going through different approaches to measure the value of consumer, even thought LOV has an obvious advantage that data collection is much simple, however, since LOV implies that consumption decision are not influenced by a consumer's concern for the welfare of others or conformity to social norms and apparently these implications are not what happen in real life.

2.4 People oriented market segmentation based on lifestyle construct

The move from mass marketing to customer relationship marketing requires decision-makers to come up with specific strategies for each individual customer based on his/her profile (Shaw et al., 2001). In today's environment of complex and ever changing customer preferences, marketing decisions that are informed by knowledge about individual customers become critical (Peppers and Rogers, 1993), which means market segmentation, the first step in the marketing process, should be well-discussed and defined.

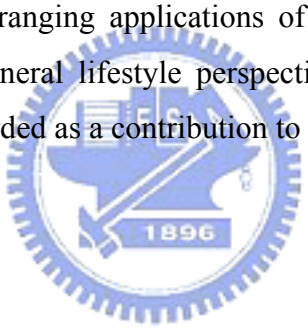
Historically, there are two different approaches to market segmentation: people oriented and product oriented (Plummer, 1974). Product oriented segmentation has been applied commonly to better understand the structure of the market of a specific product, directly or

indirectly through consumers. This approach has been more extensively used in academic research, for instance, wire-related lifestyle, wine-related lifestyle, food-related lifestyle and travel-related lifestyle which are developed to better explain the consumer behavior of specific clusters divided by lifestyle variables toward specific product or event. However, this approach is still adequate in its description of the consumer as a person and the analytic results usually do not extend to other products or services consumed by individuals since their lifestyles are not truly being measured. It is important to note that the product-specific approach has some disadvantages, one is when carried close to its extreme, it degenerates into simple redundancy, even when the most discriminating product-related variables are not that redundant, and they are often obvious translations of what the product can or cannot do. Another disadvantage of the product specific approach is that each product requires a separate study which is expensive and time-consuming and since the variables in one product-oriented study will necessarily be different from the variables in another, cross-checking and verification from study to study are foreclosed, which means when the object-specific approach is employed, each study is necessarily a "rather ad hoc and isolated exercise, requiring repetition each time a new problem arises" (Wind and Green, 1974) People oriented segmentation, usually based upon such factors as geographic, demographic, socio-economic and psycho-graphic characteristics (Be´cherel, 1999; Boote, 1981; Dolnicar, 2004; Gunter & Furnham, 1992; Swarbrooke & Horner, 1999), has the advantages to widespread usage and tends to have better explanation power to describe customers and will allow more extending usages in different markets.

There are various data mining techniques to assist to analyze and predict purchasing behavior, that is, statistical tools that help labeling or categorizing a set of cases in a database into different classes so companies can develop marketing strategies according the analytic results specifically. But when facing the tremendous product categories and customer differences, it is still a difficult task to develop accurate marketing strategies for one to one marketing in reality, especially the data collected mainly based on demographics in customer profiling. Among different constructs used to segment the market, according to Plummer (1974), demographics and socio class have received board acceptance because of the ease of quantification and consumer classification, however, both of them lack richness and need to be supplemented with other data to obtain meaningful insights into customers; psychological characteristic, provide richer information but may lack reliability when applied to mass consumers and have problems to implement. Lifestyle, combines the excellence of demographics with the richness and dimensionality of psychological

characteristics and depth research, has a longstanding history in marketing research. Lifestyle construct was first introduced by Lazer (1964), was mainly used as an umbrella term for arbitrary assortments of "activities, interests and opinions" items (AIO; Wells and Tigert, 1971) by which marketing researchers sought to describe how consumer segments differed from another. Until now, Consumer lifestyle-based segmentation has been successfully used to profile and predict the consumer market segments of a number of products and services (Fournier et al., 1992; Orth et al., 2004). Lifestyle, social and family-related variables are found to have a greater ability to profile market segments and explain segment membership than demographic variables (Koivumaa-Honkanen et al., 2004).

In this paper, as described before, are going to applied the people oriented market segmentation based on general lifestyle construct, since market segmentation by people oriented segmentation would allow more in-depth awareness of variables influencing consumers' behavior, regardless of the product or service consumed. This would bring with it the possibility of more wide-ranging applications of the methodology (Gonzalez and Laurentino, 2002). Since the general lifestyle perspective has not been widely applied, choosing the method can be regarded as a contribution to this research.



Chapter 3 Research Methods

3.1 Research Design

To solve the problem of marketing segmentation of lifestyle, a research structure is designed (see 3.1.1). Additionally, research process (see 3.1.2) is also designed according to research structure to support the operation.

3.1.1 Research Structure

The structure of this research is shown as Fig.4.

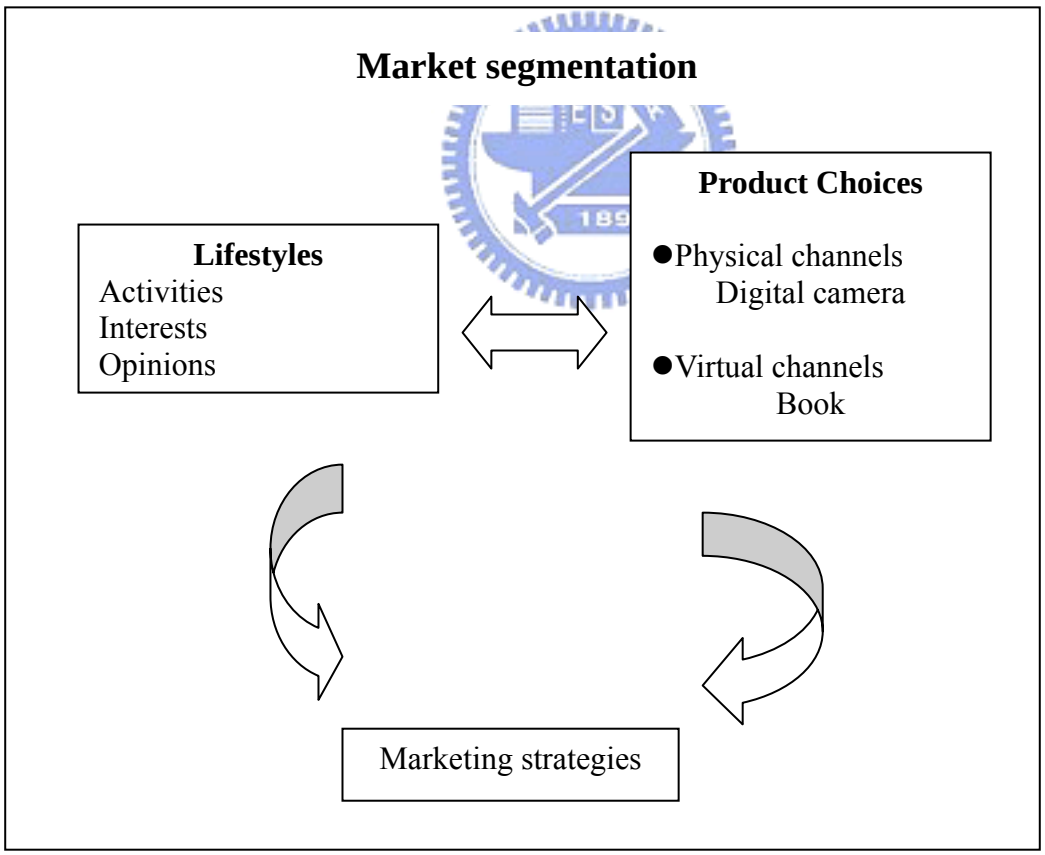


Fig. 4 Research structure

According to the motivation and objectives in this research, the research structure is proposed in Fig. 5. There are three main elements of the survey variables, that is, personality, lifestyle and demographics which are going to be extracted from the existing

literature through expert advice and literature review and will be applied as the basis of market segmentation. To improve the efficiency in data collection, two types of scales are combined to form the measurement structure in the questionnaire: ordinal and nominal. A general survey will be implemented after the questionnaire is formed and the data collected will be analyzed by data mining tools such as rough set and classification and regression tree to discover the rules of consumer choices toward different products. And not only to generate rules but also to compare the results from different segmentation tools, also, the relations between different variables will be discussed according to the analysis results. After explaining the analysis results, marketing strategies and recommendation for different products will be establish.

3.1.2 Process of Analysis and a research diagram

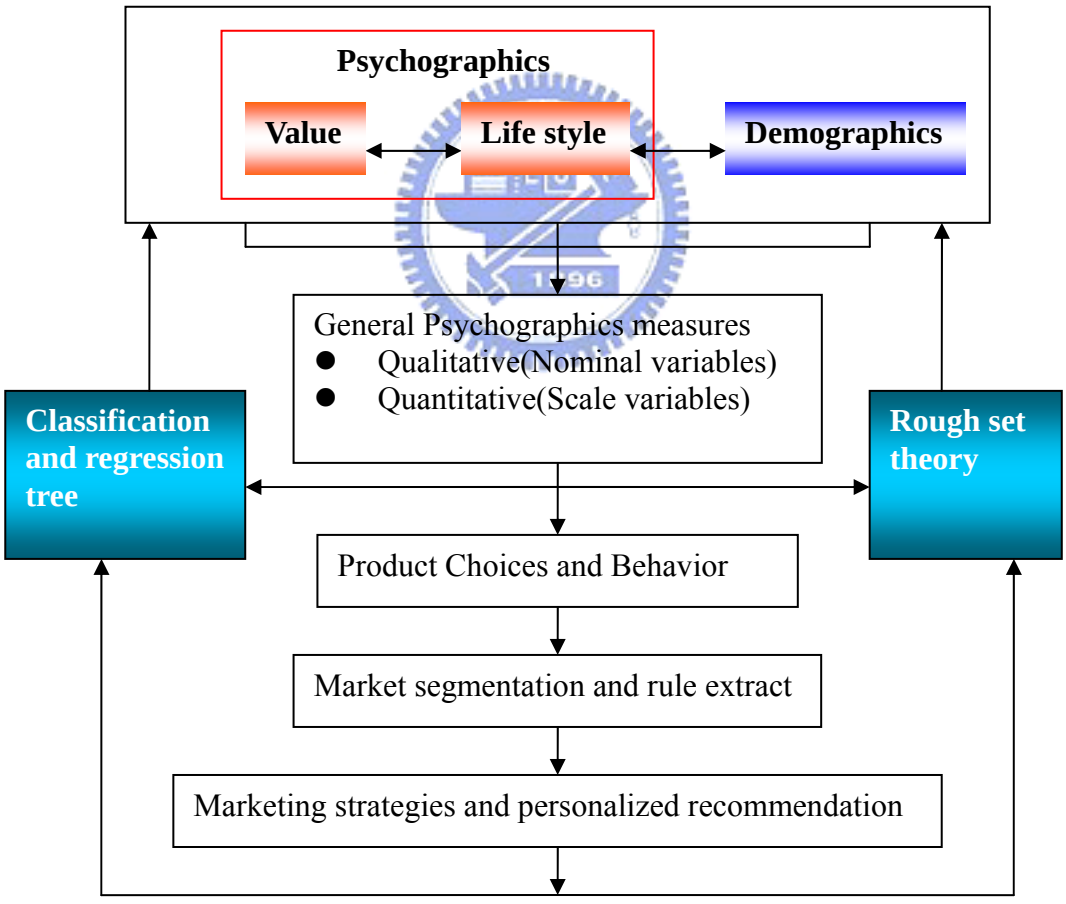


Fig. 5 Research diagram

The main purpose of using two methodologies in the research is to integrated CART and rough sets so that the ability of marketing segmentation in CART can be retrained and the

accuracy of CART can be improved by using rough sets. The procedures of the proposed method can be described as follows.

Similar to the first step of knowledge discovery from database (KDD), raw data are first preprocessed so that noise and inconsistent data are removed. Next, processed data are inputted to CART to obtain the result of marketing segmentation. Then, if some nodes are unsatisfied, i.e., the classification accuracy of a node is not satisfied by the decision-maker, the data within these nodes are retrained by rough sets and be considered as a particular segmentation. Finally, from the results of CART and rough sets, we can determine the appropriate number of segmentation and the corresponding accuracy.

The procedure of the proposed method can be depicted by Fig. 6:

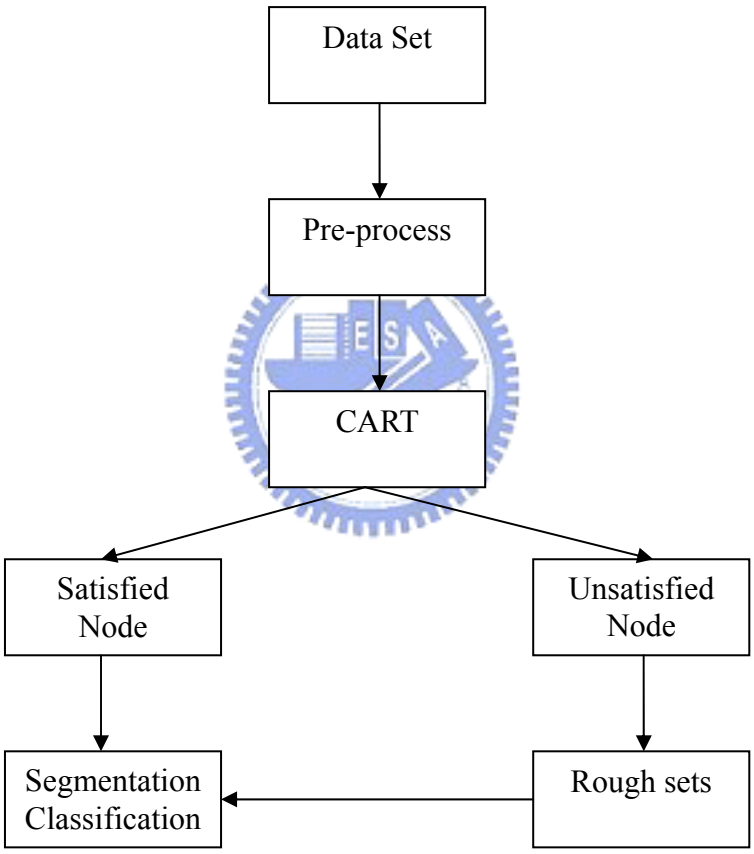


Fig. 6 The procedures of the proposed method

Then, the research process of the dissertation can be depicted as shown in Fig. 7.

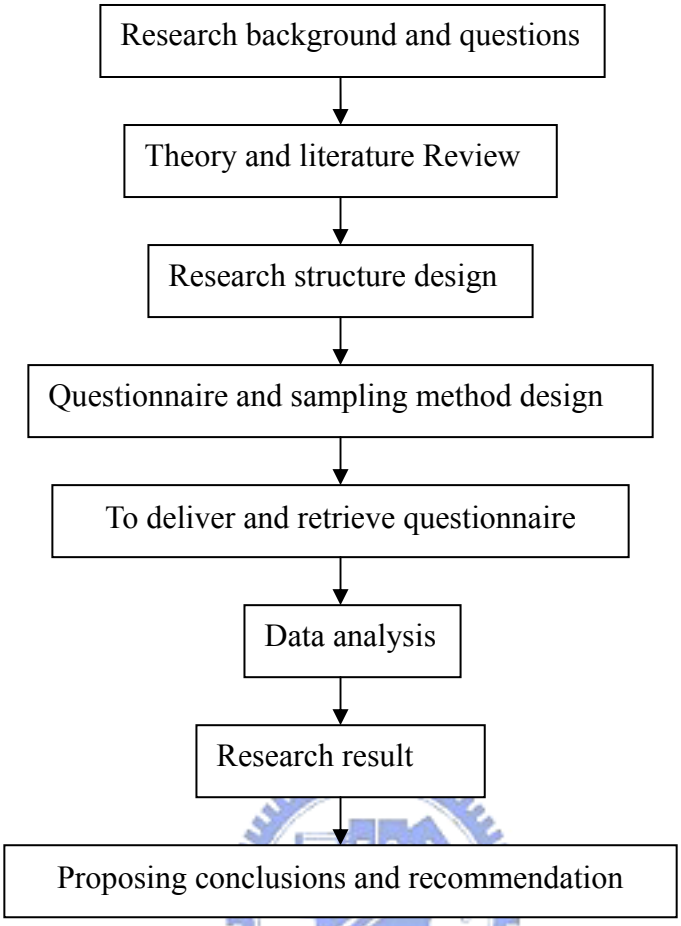


Fig. 7 Research process

The process of this study is as followed:

1. Define the research questions and objectives

   The first step is to define the research questions and objectives as a goal and criteria to operate.

2. Theory and literature review

   According to the research questions and objectives, related theories and literature which were proposed by other researchers in academic world is going to be reviewed, arranged and discussed. The result of it will be an important essence of the conceptual structure in this study.

3. Research structure design

   Refer to the theories, principals, experimental rules and the results of other researches, the appropriate structure is established to match the research propose.

4. Questionnaire and sampling method design

Questionnaire design will be generated based on the systematic arrangement of the existing academic papers and conceptual structure of this research. After considering the limitation of time, human resource, and finance, the applicable sampling method will be designed to this research.

5. To deliver and retrieve questionnaire

   According to research scope and adequate ways to deliver the questionnaires to the research object, after collecting enough number of samples, then we can start to analyze the data.

6. Data analysis

   According to the data collected in the last process, data-mining tools such as rough set and classification and regression tree are going to applied to deal with the data.

7. Research result

   Research results will be generated after the data analysis process.

8. Proposing conclusions and recommendation

   The conclusions and suggestions will be described based on the research results; the recommendation of marketing strategies will also be addressed in this process.

3.2 Analytical Methods

In this study, classification and regression tree and rough set theory are going to be applied to extract the rules of consumer choices.

3.2.1 Classification and Regression tree (CART)

Classification and regression trees (CART) is one of the methods of decision trees. Decision trees is an efficient data-mining technique which belongs to induction algorithm and machining learning. It can classify data in huge quantity according to the input variables and present the process as a diagram. Usually this kind of classify can also help to discover the lifestyle variables and build the clear hierarchical structure of the data. There are also several advantages of decision trees state as follows:

1. The rules generated can be easily understood. It means that the rules obtained by this technique can be transfer into simple English or SQL language; it's also the most powerful strength of this technique;
2. Decision trees is a good tool in the rule-oriented field;

3. It can save the time of calculation;
4. Classification trees can be computed for continuous predictors, categorical predictors or a mix of the two types;
5. It points out the ability of the best variable. (Root node is the best classifier).

3.2.1.1 History of development

CART is a binary decision trees technique, proposed by Breiman et al. (1984) which uses partitioning techniques to classify observations as binary and sequential fashion methods to increase the ability of prediction. It's a technique that can select from among a large number of variables those and their interactions that are most important in determining the outcome variable to be explained (Yohannes, 1999). It is primarily used to form prediction rules for an outcome variable based on the values of predictor variables. Even though a variety of traditional statistical approaches can be used to predict the classification of cases from complex data sets, CART analysis has been cited repeatedly as a powerful non-parameter approach in applied fields where classification or prediction are of concern, such as medicine and mental health (Johnson et al., 2002). Also, its graphical features were cited as a primary advantage over other methods. Therefore, after considering the characteristic and ability of CART, apply it to market segmentation can be appropriate.

CART has been applied in many different fields. At the very beginning, it is applied to medical diagnosis and prediction (Goldman et al., 1998; Mair et al., 1995; Thomssen et al., 1998) and mental health (Barnes et al., 1991; Boerstler et al., 1991; Craig et al., 1997). In finance field, Kao& Shumaker (1999) discovered the relationship between macroeconomic variables and the change of two different indexes return rate, the results showed that CART can provide 74% correct rate. Sorensen et al. (2000) also applied CART to improve the rules to choose stocks, proved that CART can assist investors to pick stocks effectively into their portfolio.

3.2.1.2 Growing trees and splitting criteria

The homogeneity of nodes is defined by impurity, a measure which takes the value zero for completely homogeneous nodes, and increases as homogeneity decreases. Thus maximizing the homogeneity of the groups is equivalent to minimizing their impurity.
Many measures of impurity (splitting criteria) exist, and enable us to analyze many types of responses. There are five commonly used measures (Breiman et al., 1984); three for

classification trees and two for regression trees.

For classification trees, impurity is defined in terms the proportions, c, of responses in each category. The three common criteria (indices) are:

(1) The information (entropy) index takes the form $-\sum c\ln(c)$, where $\sum$ indicates summation over categories. This index is identical to the Shannon-Weiner diversity index, and forms groups by minimizing the within-group diversity.

(2) The Gini index takes the form $1-\sum c^2$. At each split, the Gini index tends to split off the largest category into a separate group, whereas the information index tends to form groups comprising more than one category in the early splits.

(3) The twoing index can be used for more than two categories. It defines two "super categories" at each split, for which the impurity is defined by Gini index. It can also be used for ordered categories.

For regression trees, the two common forms of impurity are:

(1) Sums of squires about the group means. This is equivalent to least squares linear models.

(2) Sums of absolute deviations about the median. This gives a robust tree (Breiman et al, 1984). However, for ecological data dominated by zeros, this criterion can be ineffective, especially when the explanatory variables are categorical. In such cases all possible splits may result in groups with zero medians, and no splits will be formed.

Trees can also be formulated as statistical models, akin to linear, generalized linear and generalized additive models (Clark and Pregibon, 1992). In this approach, splits are based on an explicit statistical model, the deviance of which defines the dissimilarity measure. For classification trees, they use a multinomial model, equivalent to the information index, with the deviance defined by the multinomial log-likelihood. For regression trees, Clark and Pregible (1992) use the Gaussian model, and the deviance for a bode is simply the sums of squares about the mean. Summing over all leaves gives the overall deviance for the tree.

Pruning trees

A natural way of using a splitting criterion to grow a tree is to continue splitting until the improvement due to additional splits is less than a prespecified cutoff, and then take this as the best tree. The fundamental work of Breiman et al. (1984) points out two weakness of this approach. First, if the stopping rule is based on too small an improvement, then an overlarge tree will be result. Second, if the criterion is too large, then splits based on interactions between explanatory variables will not be discovered unless at least one of the

associated main effects is large enough to generate a split.

Breiman et al. (1984) introduce three basic ideas to solve the problem of finding the best tree. The first idea is tree pruning, rather than stop growth in progress; they grow an over-large tree and then seek ways to cut it back. This can be computationally infeasible, since the number of sub-trees is usually very large. To overcome this problem, their second idea is to find a sequence of nested trees of decreasing size, each of which is the best of all trees of its size. For this they use the resubstitution estimate of error, $R(T)$, which can be either the overall misclassification rate or the total residual ss, dependent on the type of tree. They show that, for any number $\partial(\geq 0)$ there is a unique smallest tree that minimizes $R(T)+\partial|T|$, where $|T|$ tree size (number of leaves) is. By allowing ∂ increasing form 0 to large, we obtain the desired sequence of nested trees of decreasing size, beginning with the initial overlarge tree and ending with the root tree with no splits at all. Since each tree in this sequence is the best of its size, choosing the best tree is reduced to the task of choosing the best size, a much simpler task than comparing all possible subtrees. $R(T)$ is not suitable for this choice because it will always be minimized by the largest tree (just as adding more explanatory variables reduces the residual ss of a regression). Thus, to complete the process, we require better estimates of error, and the third idea of Breiman et al. (1984) is to obtain "honest" estimates of error by cross-validation, as described in selecting tree size by cross-validation. This can be computationally demanding, but is now feasible since we only have to consider one tree of each size, i.e., the trees of the nested sequence.

Selecting trees size by cross-validation

Breiman et al. (1984) use cross-validation to obtain honest estimates of true (prediction) error for trees of a given size. For the sequence of trees, these estimates of error can be plotted against tree size, and the size with the minimum error selected. A single tree selected by cross-validation can be used for description and/or prediction. It should be interpreted as the tree which has the smallest estimated error and is the best estimated predictive single tree.

Cross-validation can be implemented in two ways. First, if enough data are available, we select a random subset of the data, typically comprising one-half to two-thirds of all data, and, using only these data, build the sequence of nested trees. For each tree, predict the response of the remaining data, and calculate the error from the predictions and the

observed values. The tree with the smallest predicted error is then selected. One drawback of this technique is that there are often insufficient data to build good trees using only a subset of the data. The second way is to use V-fold cross-validation as follows: (1) divide the data into a number, V, of mutually exclusive subsets (typically $V = 10$) of approximately equal size; (2) drop out each subset in turn, build a tree using data from the remaining subsets, and use it to predict the responses for the omitted subset; (3) calculate the estimated error for each subset (e.g., for a sums of squares regression tree, the error is the sum of squared differences of the observations and predictions), and sum over all subsets; (4) repeat steps (2)-(3) for each size of tree; and (5) select the tree with the smallest estimated error rate. The subsets can be chosen randomly, but stratification into groups according to the value of the response variable gives smaller and more accurate estimates of the true error rate (Breiman et al., 1984).

Breiman et al. (1984) suggested the 1-SE rule whereby the best tree is taken as the smallest tree such that its estimated error rate is within one standard error of the minimum. The standard error of the estimate can be calculated for each tree size. Use of the 1-SE rule can result in a much smaller tree than suggested by the minimum cross-validated-error, but within minimal increase in the estimated error rate (at most <1 Se). Irrespective of whether the minimum or 1 SE rule is used, inspection of the cross-validated sequence is necessary to ensure that the sequence of trees has been grown large enough. For both the minimum and 1-SE rule, the size of the selected tree will vary under repeated cross-validation, and it is advisable to run several cross-validations in order to assess the degree of variation in the size of the best tree, and ensure the chosen tree is not atypical.

3.2.2 Rough set theory (RST)

In this section the introduction of rough set theory and its use in analyzing the attributes of combination values for making marketing decisions. In this section the history of rough set theory is described, and in section the algorithms of the theory for decision-making are presented.

Rough set theory is used in this study to analyze the content and features of the data. The theory, which was developed by Pawlak (1982), is a rule-based decision-making technique that can handle crisp datasets and fuzzy datasets without the need for a pre-assumption membership function, which fuzzy theory requires. It can also deal with uncertain, vague, and imperceptible data. Until now, analysis of the attributes of combination values using

rough set theory has only been addressed by a few papers.

3.2.2.1 History of development

Rough set theory can deal with inexact, uncertain, and vague datasets (Walczak & Massart, 1999). Both Fuzzy Set Theory and Rough Set theory are used with the indiscernibility relation and perceptible knowledge. The major difference between them is that rough set theory does not need a membership function; thus, it can avoid pre-assumption and one-sided information analysis. A detail discussion of rough set theory can be found in Walczak and Massart (1999). Rough set theory was developed by Pawlak (1982, 1984, 2004). It ha been applied to the management of a number of the issues, including: medical diagnosis, engineering reliability, expert systems, empirical study of materials data (Jackson et al., 1996), machine diagnosis (Zhai et al., 2002), business failure prediction (Beynon & Peel, 2001; Dimitras et al., 1999), activity-based travel modeling (Witlox & Tindemans, 2004), travel demand analysis (Gon & Law, 2003), solving linear programs(Azibi &Vanderpooten, 2002), data mining (Li & Wang, 2004; Hu et al., 2003; Chan, 1998), and ∂-RST (Quafafou, 2000). Another paper discusses the preference-order of attribute criteria needed to extend the original rough set theory, such as sorting, choice and ranking problem (Greco et al., 2001). The rough set method is useful for exploring data patterns because of its ability to search through a multi-dimensional data space and determine the relative importance of each attribute with respect to its output.

Rough set theory applies the indiscernibility relation and data pattern comparison based on the concept of an information system with indiscernible data, where the data is uncertain or inconsistent. The data is grouped into classes called elementary sets. Feature/attribute selection is crucial in data processing that consists of relevant (or maybe irrelevant) object patterns, but it may be redundant in data pattern recognition. More information regarding attributes can be found in the works of Swiniarski and Skowron (2003), Polkowski (2004), and Inuiguchi (2004). The objects in a class may have a relationship with the corresponding features/attributes, and expert knowledge is used to process attribute extraction. Each elementary set is independent of the others. We can extract knowledge from each elementary set used in the real world. The details of rough set theory are presented in Appendix (Shying et al., 2007).

Chapter 4 Empirical Studies

In this part of the research, lifestyle variables and methodologies proposed in chapter 3 will be combined to develop a novel model to improve the traditional model which often uses demographics variable as the basis to segment the market. In the empirical studies, a decent survey with appropriate number of lifestyle variables are made, and here, we try to provide a new way to explain the segment found by this new model. At the end of the chapter, the combination of two methodologies is proposed to enhance the accuracy of the prediction results of this model.

4.1 Questionnaire design

In order to overcome the excessive length of lifestyle questionnaires, the variables under study were measured with two different types of scale: ordinal and nominal. The ordinal scales used were seven-point Likert scales, from strong agreement to strong disagreement. The use of variables measured on nominal scales allowed attainment of one of the main aims of this study, which was to obtain the same amount of information in considerably less time than usual with other research into lifestyle.

By cutting down on the time spent answering the questionnaire, the quality of the information gathered was improved, since it was possible to avoid the bias arising from the excessive time needed.

Moreover, it was possible to achieve a substantial reduction in the cost of field work and thus of the whole research project, saving resources and making it more affordable for companies.

The questionnaire designed for empirical analysis can be divided into different parts in this study, with the capability to deal with variables of different scales of the proposed methodologies in this study; type of scale is no longer an issue. Therefore, the questionnaire can be divided into four main sections; questions about general lifestyle of attitude and interest items are presented in the first section, in this part, two different types of questions such as Likert scale and nominal scale are both included. In the second section, questions of daily activities items are offered in nominal scale. In the third section, questions of

purchasing behavior, that is, behavior of purchasing digital camera and buying books online, are presented in nominal scale. Demographic variables are in the last section of the questionnaire.

Table 3 The variables used in this research

First block: Lifestyle <Interest and opinions>	Type of variable: ordinal and nominal variable
Questions about ◆Family ◆Society ◆Politics ◆Personal success factors ◆Attitude to personal problems ◆Environment ◆Technology ◆Fashion ◆Health ◆Finance and saving	
Second block: Lifestyle <Leisured activities>	Type of variable: nominal variable
Questions about ◆Outdoor activities ◆Sports ◆Art and Learning activities ◆Indoor activities	
Third block: Purchasing behavior	Type of variable: nominal variable
Questions about ◆Digital camera ◆books online	
Fourth block: Demographics	Type of variable: ordinal and nominal variable
Questions about ◆Gender ◆Age ◆Education ◆Profession ◆Marital status ◆Income ◆District of Residence	

4.2 Sample descriptions

Data for analysis in this study is collected by distributing the self-completion questionnaires. The period of distributing questionnaires was from January 1 to March 31 in 2009. In the questionnaire conventional AIO type questions and demographic variables were both included. Due to the time limitation and financial constrains, convenience sampling was utilized to collect the data in this study. A total of 525 questionnaires were collected from universities and companies. However, due to incomplete responses on some questionnaires, 367 questionnaires were accepted for the final sample and used for data analysis. The demographic characteristics of the participants are presented in Table 4. From Table 4, it can be seen that the sample is 48.1% male and 51.9% female, most respondents

are around 20 to 34 years old, 54.3% are college students and 34.4% are working as full time employees. Most of them did not get married; 77.4% salaries are below 30,000 NT per month and live in the north part of Taiwan (79.3%).

Table 4 Demographic characteristics of the sample

Demographic characteristics	**category**	**Percentage (%)**
Gender	Male	48.1
	Female	51.9
Age	Under 20 years	26.9
	20-24 years	44.0
	25-29 years	16.6
	30-34 years	5.5
	35-39 years	2.8
	40-44 years	2.2
	45-49 years	0.8
	50 years and over	1.1
Education	Primary and jour high school	0.3
	Senior high school	4.8
	Bachelor degree	74.4
	Postgraduate (Masters or Ph.D.)	20.2
Profession	Full time job	34.4
	Part time job	9.9
	Housewife	0.6
	Retire	0.3
	School	54.3
	Unemployed	0.6
Marital status	Never married	97.1
	Married	2.9
Income (per	0-10000 NT	53.8

month)	10001-20000 NT	10.8
	20001-30000 NT	12.8
	30001-40000 NT	6.7
	40001-50000 NT	6.1
	50001-60000 NT	3.5
	60001-70000 NT	1.7
	70001-80000 NT	1.5
	80001-90000NT	0.6
	90001-100000NT	0.6
	More than 100000 NT	2.0
District of residence	North	79.3
	Center	11.2
	South	8.7
	East	0.8
Purchase digital camera	Yes	47.0
	No	53.0
Purchase books online	Yes	58.8
	No	41.2

4.3 The results of the empirical analysis

According to the questionnaire design, general lifestyle variables include activity, interest and opinion items with respect to Likert and nominal scales. To deal with different types of variables by CART, the combinations of different items and scales are designed to discover the important lifestyle variables to various purchasing behavior. Therefore, seven models are generated by different combinations of lifestyle items and variable scales for the behavior of purchasing digital camera and books online, respectively. Model 1 presents the interest and opinion items with Likert scale, model 2 presents the interest and opinion items with nominal scale, model 3 presents the activities items with nominal scale, model 4

presents the combination of the interest and opinion items with both Likert and nominal scale, model 5 presents the combination of the interest and opinion items with nominal scale and activity items with nominal scale, model 6 presents the combination of the interest and opinion items with Likert scale and activity items with nominal scale, model 7 presents the combination of the interest and opinion items with Likert and nominal scale and activity items with nominal scale.

In the process of running CART, split sample validation with random assignment is selected, which means samples are separated into two parts, the first part of the sample is used to train the initial classification tree and the other part is used to exam the classification accuracy of the initial tree. In this study, 70% of total samples are used to train the classification tree and 30% are used to test the training result , since the software selects those training and testing sample randomly, it may result in different variables extracted and accuracy, hence, 10 times of trials of each model will be shown here.

In the following content, there are two main experimental analyses to present; first, the analytical results will be presented by applying the novel model proposed in this study, that is, the combination of lifestyle variables and CART. The improved model which integrated two methodologies will be proposed and the results will be shown in the second part.

4.3.1 Empirical results – Digital Camera

In the model 1, 58 general lifestyle variables in Likert scale are the input variables, after computing process of CART, the result of 10 times trials are shown in Table 5. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V29, V39), (V8, V40), (V52), (V46, V12) are effective explanatory variables which can used to predict the purchasing behavior of digital camera, the details of the result of those segmentation variables are shown in Table 6. Among those extracted variables, (V46, V12) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 8. From the tree, it can be explained that people who show more willingness to by high-tech product even if they are not familiar with and willing to pay more on clothes to dress up themselves have more possibilities to buy a digital camera.

Table 5 Experimental results of digital camera purchasing behavior using ordinal lifestyle variable by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V29, V39)**	59.0	52.5
2	(V8)	56.9	43.9
3	**(V8, V40)**	60.3	53.3
4	(V33)	61.1	43.6
5	**(V52)**	58.3	51.8
6	**(V52)**	58.3	51.8
7	**(V52)**	59.0	50.5
8	**(V46, V12)***	61.4	57.5
9	(V33)	55.9	48.1
10	(V25, V7)	63.9	50.0

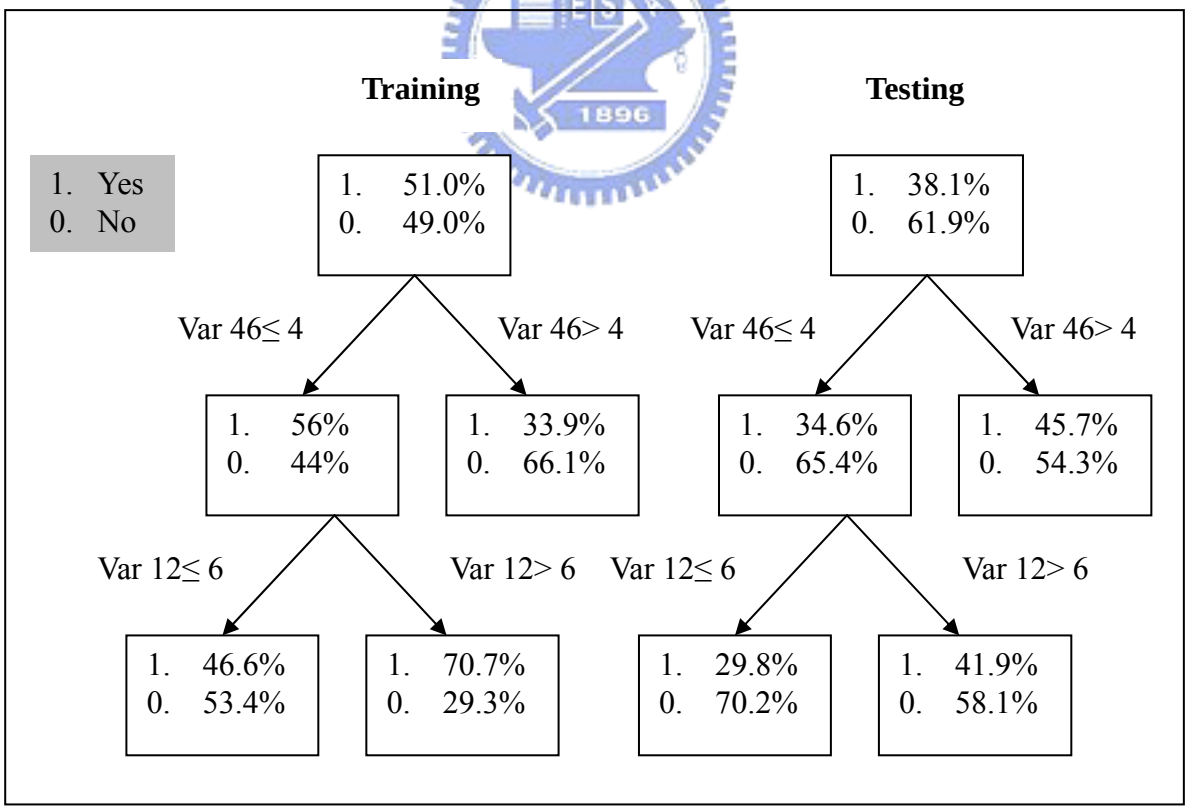


Fig. 8 The marketing segmentation of digital camera by CART-the best trial of model 1

Table 6 The result of market segmentation of digital camera by CART- model 1

Model 1	Variables extracted by CART and Classification results	Willingness to buy a DC
(V29, V39)	V29: I trust people easily V39: I am willing to buy the product which protects the earth	
	V29>6	Yes
	V29≤6 + V39≤4	No
	V29≤6 + V39≥4	No
(V8, V40)	V8: I like to spent much time on housework V40: I often feel hesitant when making decisions	
	V8>4	No
	V8≤4 + V40≤6	No
	V8≤4 + V40>6	Yes
(V52)	V52: When facing minute and complicated product functions, I will try to overcome those difficulties	
	V52≤4	Yes
	V52>4	No
(V46, V12)	V46: I am willing to try and purchase unfamiliar technological products V12: I am willing to spend money on dressing myself	
	V46>4	No
	V46≤4 + V12≤6	No
	V46≤4 + V12>6	Yes

In the model 2, 10 general lifestyle variables in nominal scale are the input variables. After computing process of CART, the results of 10 times trials are shown in Table 7. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V64) and (V62, V64) are effective explanatory variables which can use to predict the purchasing

behavior of digital camera, the details of the result of those segmentation variables are shown in Table 8. Among those extracted variables, (V62, V64) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig.9. From the tree, it can be explained that people who spent 5 to 15 hours or more than 26 hours online and live in the house or mansion have more possibilities to buy a digital camera.

Table 7 Experimental results of digital camera purchasing behavior using nominal lifestyle variables by CART

Trial	**Variable selection**	**Accuracy of classification (Training) (%)**	**Accuracy of classification (Testing) (%)**
1	**(V64)**	56.4	50.0
2	(V64, V67, V61)	60.2	47.2
3	**(V64)**	58.8	52.7
4	(V64)	59.1	43.6
5	**(V64)**	55.2	52.6
6	**(V62)**	57.1	50.0
7	**(V64)**	59.0	51.5
8	(V61)	54.8	49.1
9	(V61)	55.9	44.4
10	**(V62, V64)***	59.5	55.5

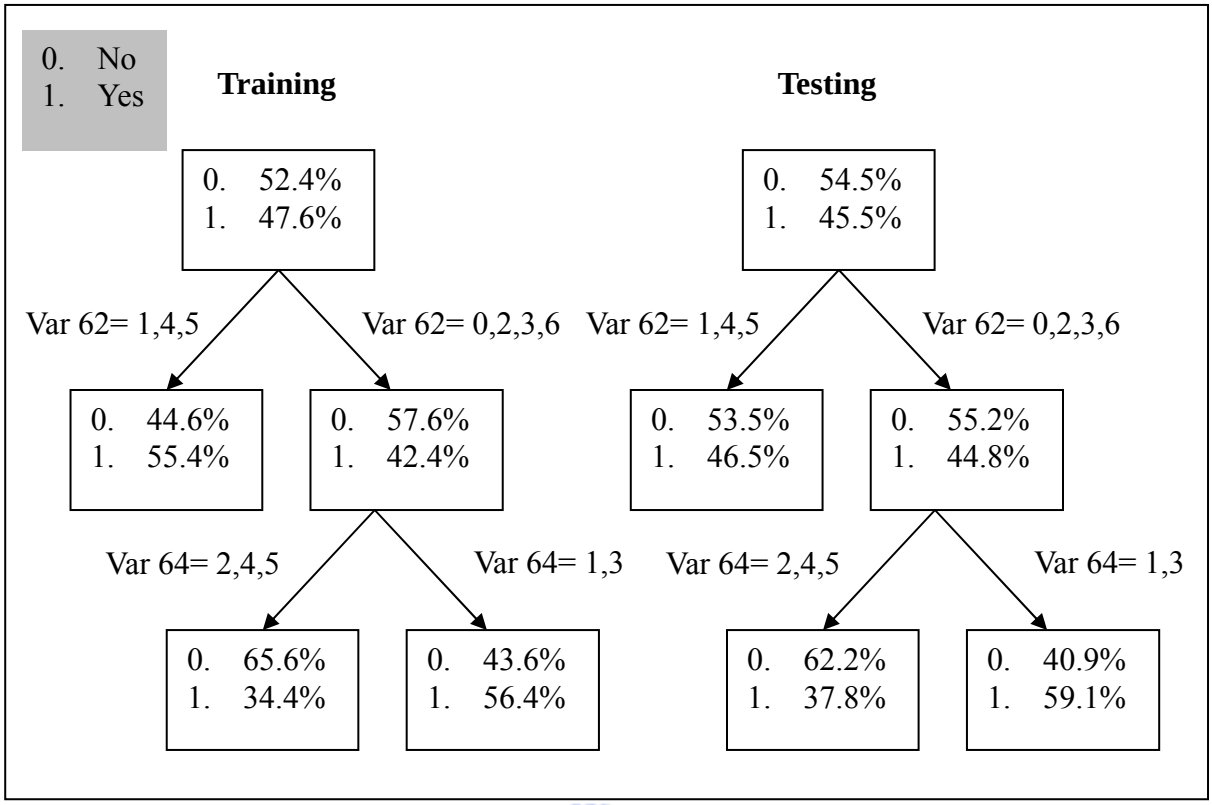


Fig. 9 The marketing segmentation of digital camera by CART-the best trial of model 2

Table 8 The result of market segmentation of digital camera by CART- model 2

Model 2	**Variables extracted by CART and Classification results**	**Willingness to buy a DC**
(V64)	V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others	
	V64=4	No
	V64=2,1,5,3	Yes
(V62, V64)	V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others	
	V62=4, 1, 5	Yes
	V62=2,3,6,0 + V64=2,4,5	No
	V62≤4 + V64=1,3	Yes

In the model 3, 52 general lifestyle variables in nominal scale are the input variables, after computing process of CART, the result of 10 times trials are shown in Table 9. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V89), (V104) and (V80) are effective explanatory variables which can use to predict the purchasing behavior of digital camera, the details of the result of those segmentation variables are shown in Table 10. Among those extracted variables, (V104) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig.10. From the tree, it can be explained that people who are engaged in photographic activities have more possibilities to buy a digital camera.

Table 9 Experimental results of digital camera purchasing behavior using nominal activities variables by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V89)**	57.0	52.1
2	**(V104)**	59.4	56.4
3	**(V104)**	59.3	56.6
4	**(V104)***	58.1	59.3
5	**(V104)**	59.6	55.9
6	**(V104)**	60.9	52.8
7	(V74, V80)	58.8	46.5
8	(V104)	63.2	49.6
9	(V104, V71)	58.6	46.7
10	**(V80)**	57.0	50.5

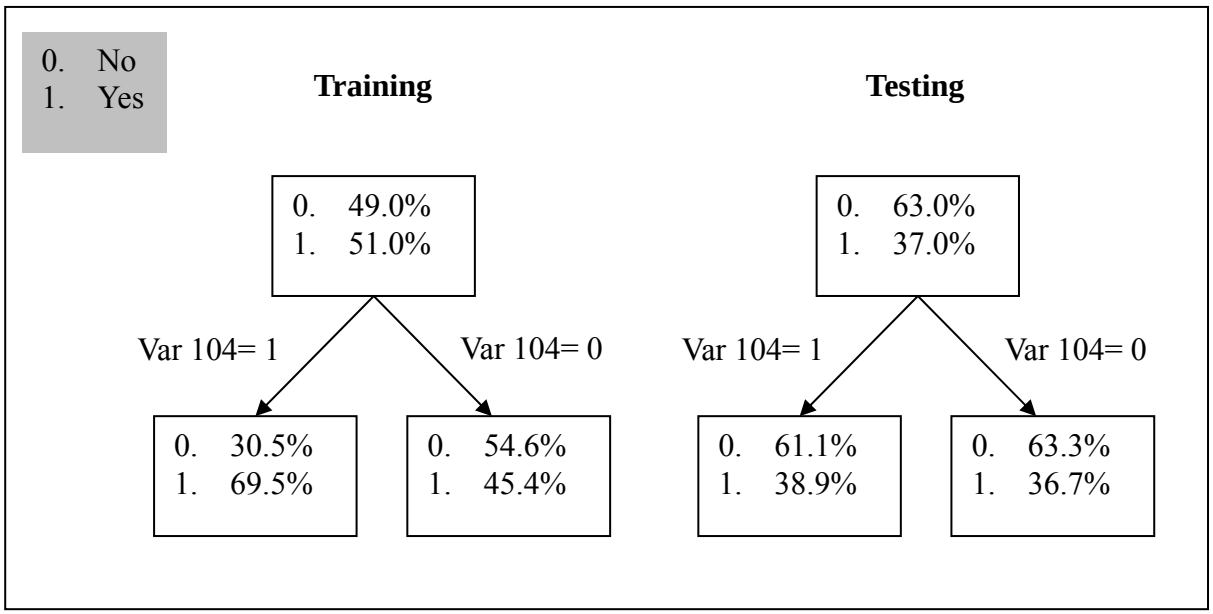


Fig. 10 The marketing segmentation of digital camera by CART-the best trial of model 3

Table 10 The result of market segmentation of digital camera by CART- model 3

Model 3	**Variables extracted by CART and Classification results**	**Willingness to buy a DC**
(V89)	V89: Swimming 0: No; 1: yes	
	V89=0	No
	V89=1	Yes
(V104)	V104: Photography 0: No; 1: yes	
	V104=0	No
	V104=1	Yes
(V80)	V80: Like to eat in the restaurant 0: No; 1: yes	
	V80=0	No
	V80=1	Yes

In the model 4, 58 general lifestyle variables in Likert scale plus 10 general lifestyle variables in nominal scale are the input variables, after computing process of CART, the

result of 10 times trials are shown in Table 11. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V64), and (V64, V33) are effective explanatory variables which can use to predict the purchasing behavior of digital camera, the details of the result of those segmentation variables are shown in Table 12. Among those extracted variables, (V64) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig.11. From the tree, it can be explained that people who don't live in a suite have more possibilities to buy a digital camera.

Table 11 Experimental results of digital camera purchasing behavior using ordinal and nominal lifestyle variables by CART

Trial	**Variable selection**	**Accuracy of classification (Training) (%)**	**Accuracy of classification (Testing) (%)**
1	**(V64)**	56.4	50.0
2	(V64, V31)	64.7	44.5
3	**(V64)***	55.2	52.6
4	**(V64)**	59.0	51.5
5	(V64)	59.2	43.8
6	(V7)	58.1	44.0
7	(V33)	57.9	42.7
8	**(V64, V33)**	62.1	50.5
9	(V19)	57.6	46.7
10	(V64)	60.1	49.0

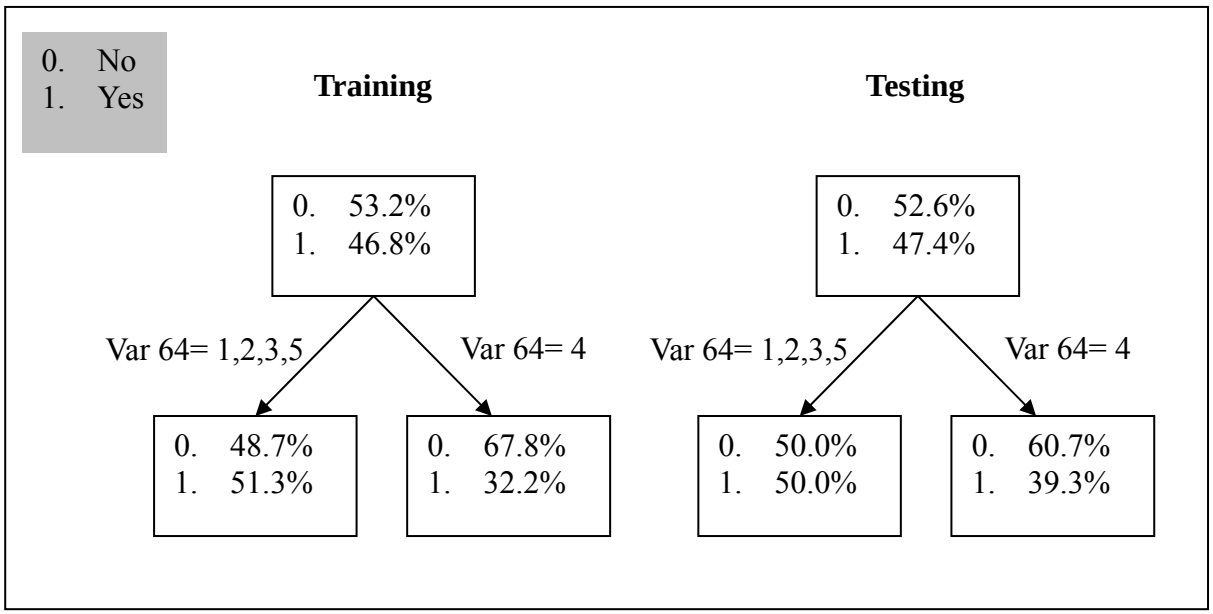


Fig. 11 The marketing segmentation of digital camera by CART-the best trial of model 4

Table 12 The result of market segmentation of digital camera by CART- model 4

Model 4	**Variables extracted by CART and Classification results**	**Willingness to buy a DC**
(V64)	V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others	
	V64=4	No
	V64=2,1,5,3	Yes
(V64, V33)	V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others V33: I will compare the price carefully when buying stuff	
	V64=4	No
	V64=2,1,5,3 + V33≤4	No
	V64=2,1,5,3 + V33>4	Yes

In the model 5, 10 general lifestyle variables (interest and opinion)in nominal scale plus 52 general lifestyle variables (activities) are the input variables, after computing process of CART, the result of 10 times trials are shown in Table 13. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of

purchasing behavior. Therefore, from the result of CART, (V104), (V104, V87, V64), (V64, V73), and (V104, V64) are effective explanatory variables which can use to predict the purchasing behavior of digital camera, the details of the result of those segmentation variables are shown in Table 14. Among those extracted variables, (V104, V87, V64) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 12. From the tree, it can be explained that people who are engaged in photographic activities or people who are not engaged in photographic and jogging activities but live in the house, mansion or others have more possibilities to buy a digital camera.

Table 13 Experimental results of digital camera purchasing behavior using nominal lifestyle and nominal activities variables by CART

Trial	**Variable selection**	**Accuracy of classification (Training) (%)**	**Accuracy of classification (Testing) (%)**
1	**(V104)**	59.7	55.3
2	**(V104, V87, V64)***	61.8	61.1
3	**(V64, V73)**	61.8	56.2
4	**(V104)**	58.3	58.2
5	**(V104,V64)**	61.4	55.0
6	**(V104)**	59.0	56.4
7	**(V104)**	58.7	57.3
8	(V64)	57.5	47.3
9	**(V104)**	59.4	55.7
10	**(V64, V73)**	63.1	51.8

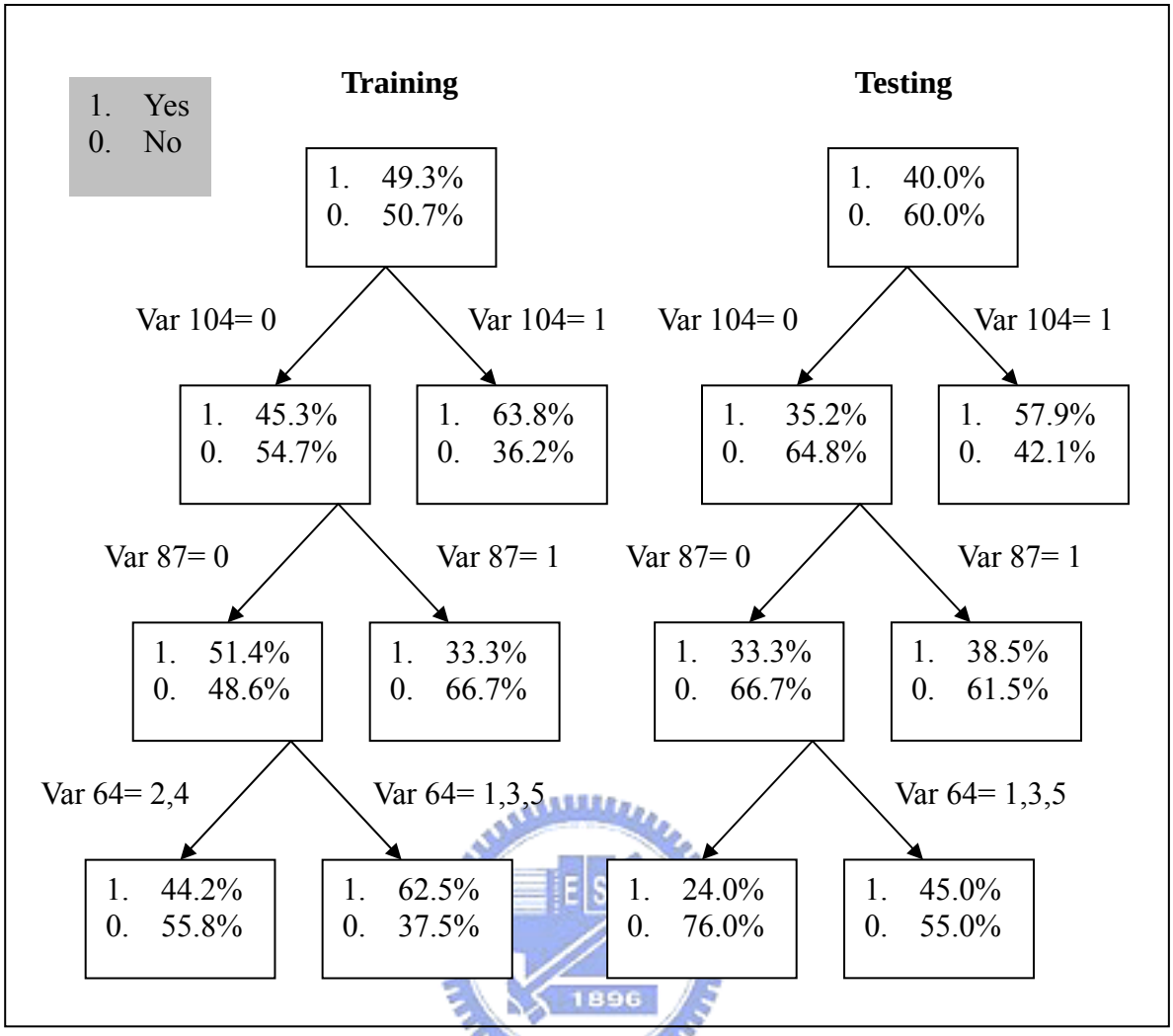


Fig. 12 The marketing segmentation of digital camera by CART-the best trial of model 5

Table 14 The result of market segmentation of digital camera by CART- model 5

Model 5	**Variables extracted by CART and Classification results**	**Willingness to buy a DC**
(V104)	V104: Photography 0: No; 1: yes	
	V104=0 V104=1	No Yes
(V104, V87, V64)	V104: Photography 0: No; 1: yes V87: Jogging 0: No; 1: yes	

	V64: House types 1: House 2: Apartment 3. Mansion 4. Suite 5. Others	
	V104=1 V104=0 + V87=1 V104=0 + V87=0 +V64=2,4 V104=0 + V87=0 +V64=3,1,5	Yes No No Yes
(V64, V73)	V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others V73: Cinema 0: No; 1: yes	
	V64=4 V64=2,1,5,3 + V73=0 V64=2,1,5,3 + V73=1	No No Yes
(V104, V64)	V104: Photography 0: No; 1: yes V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others	
	V104=1 V104=0 + V64=2,4 V104=0 + V64=3,5,1	Yes No Yes

In the model 6, 58 general lifestyle variables (interest and opinion) in Likert scale plus 52 general lifestyle variables (activities) are the input variables. After computing process of CART, the results of 10 times trials are shown in Table 15. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V104) is the only effective explanatory variables which can use to predict the purchasing behavior of digital camera; the details of the result of those segmentation variables are shown in Table 16. And the training and

testing tree of this trial is shown in Fig.13. From the tree, it can be explained that people who are engaged in photographic activities have more possibilities to buy a digital camera.

Table 15 Experimental results of digital camera purchasing behavior using nominal lifestyle and nominal activities variables by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	(V104, V40)	61.8	45.1
2	(V52)	61.8	44.8
3	**(V104)***	57.7	59.6
4	**(V104)**	59.4	55.9
5	**(V104)**	60.6	52.8
6	(V3, V18)	60.5	34.3
7	**(V104)**	59.0	57.5
8	**(V104)**	59.8	55.2
9	(V8)	59.6	43.9
10	(V52)	62.3	42.7

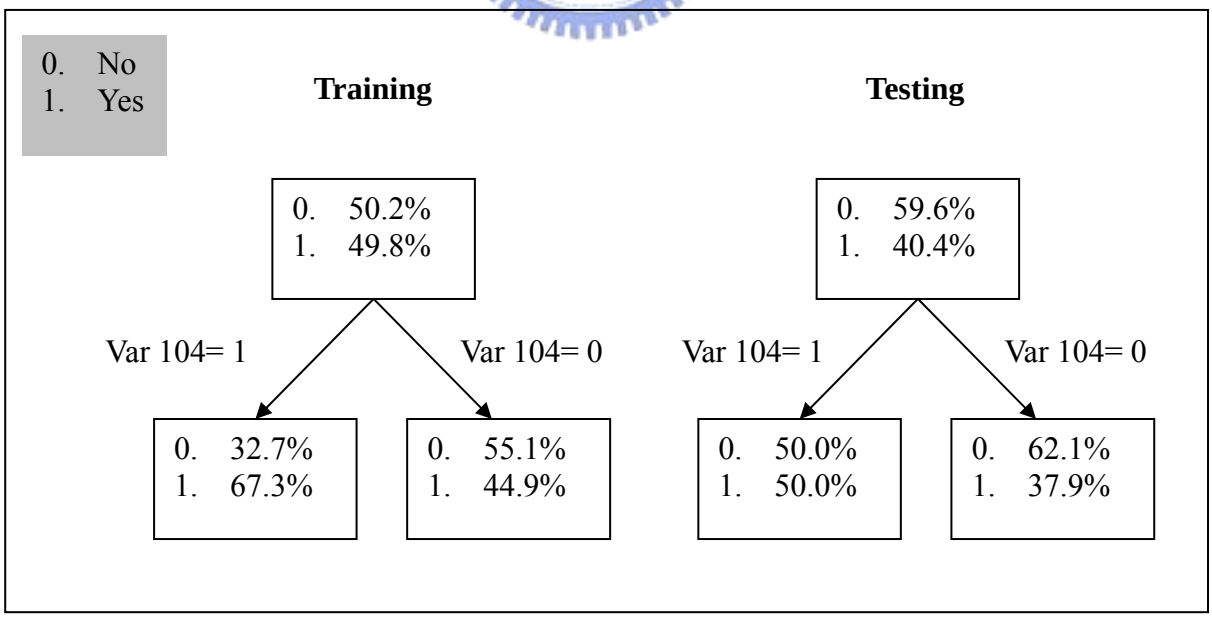


Fig. 13 The marketing segmentation of digital camera by CART-the best trial of model 6

Table 16 The result of market segmentation of digital camera by CART - model 6

Model 6	Variables extracted by CART and Classification results	Willingness

		to buy a DC
(V104)	V104: Photography 0: No; 1: yes	
	V104=0 V104=1	No Yes

In the model 7, 58 general lifestyle variables (interest and opinion) in Likert scale, 10 general lifestyle variables (interest and opinion) in nominal scale and 52 general lifestyle variables (activities) are the input variables. After computing process of CART, the results of 10 times trials are shown in Table 17. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V104) and (V29) are effective explanatory variables which can use to predict the purchasing behavior of digital camera, the details of the result of those segmentation variables are shown in Table 18. Between these two extracted variables, (V104) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 14. From the tree, it can be explained that people who are engaged in photographic activities have more possibilities to buy a digital camera.

Table 17 Experimental results of digital camera purchasing behavior using ordinal and nominal lifestyle and nominal activities variables by CART

Trial	**Variable selection**	**Accuracy of classification (Training) (%)**	**Accuracy of classification (Testing) (%)**
1	**(V104)**	60.6	52.4
2	(V52, V28)	59.9	49.2
3	(V64, V47)	61.3	48.5
4	**(V104)**	60.3	53.6
5	**(V104)***	60.2	53.8
6	(V29, V94)	60.5	49.0
7	(V47)	55.6	47.4

8	(V104)	62.2	48.0
9	(V80)	58.8	48.4
10	**(V29)**	59.5	52.5

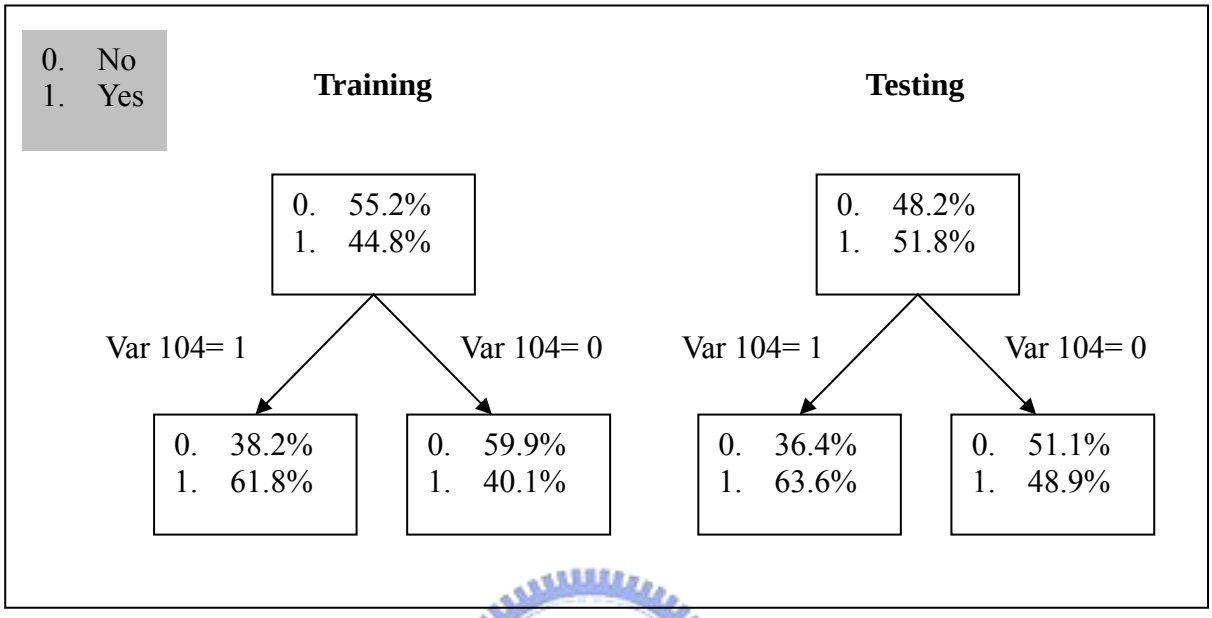


Fig. 14 The marketing segmentation of digital camera by CART-the best trial of model 7

Table 18 The result of market segmentation of digital camera by CART- model 7

Model 7	**Variables extracted by CART and Classification results**	**Willingness to buy a DC**
(V104)	V104: Photography 0: No; 1: yes	
	V104=0 V104=1	No Yes
(V29)	V29: I trust people easily	
	V29≤6 V29>6	No Yes

Table 19 Experimental results of digital camera purchasing behavior using demographics variables by CART

Trial	**Variable selection**	**Accuracy of classification (Training) (%)**	**Accuracy of classification (Testing) (%)**

1	**(income, profession)**	60.2	51.9
2	**(gender)**	57.5	55.0
3	**(gender, income)**	58.0	53.7
4	**(income)**	58.2	53.3
5	**(income, profession)**	62.1	51.6
6	**(income, gender)**	61.3	56.8
7	**(income, gender)**	59.0	51.8
8	(income)	56.3	48.6
9	**(gender)**	56.6	56.1
10	**(income, gender)**	59.1	52.9

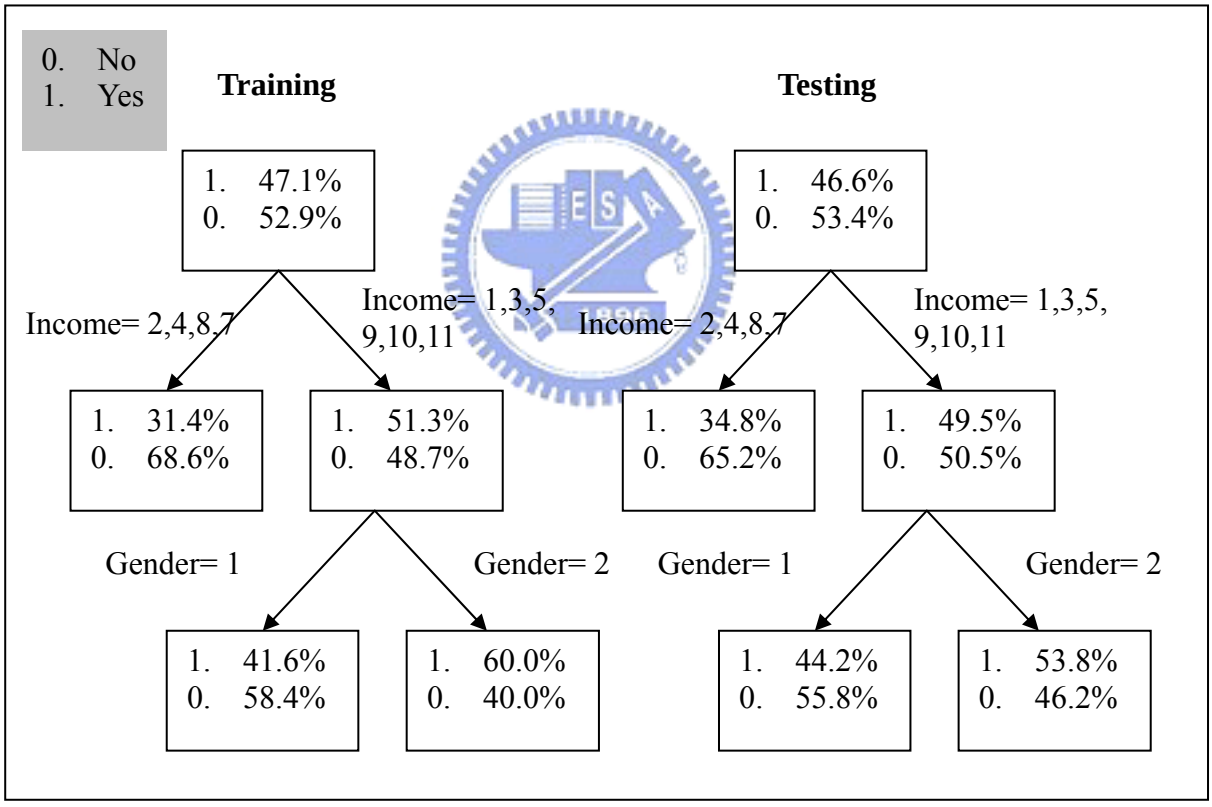


Fig. 15 The marketing segmentation of digital camera by CART-the best trial of model 8

Table 20 Summary of lifestyle variable selection of digital camera from different models by CART

model	**Input variable**	**Number of variable**	**Total variable selected from the model**	**Accuracy of classification of the best trial (Training) (%)**	**Accuracy of classification of the best trial (Testing) (%)**

1	Lifestyle(Scale)	58	V29,V39,V52, V8,V40,V46, V12	61.4	57.5
2	Lifestyle(Nominal)	10	V64, V62	59.5	55.5
3	Activities(Nominal)	52	V89, V104, V80	58.1	59.3
4	Lifestyle (Scale+ Nominal)	68	V64,V33	55.2	52.6
5	Lifestyle(Nominal)+ Activities(Nominal)	62	V104,V87, V64, V73	61.8	61.1
6	Lifestyle(Scale)+ Activities(Nominal)	110	V104	57.7	59.6
7	Lifestyle(Scale, Nominal)+ Activities(Nominal)	120	V104,V29	60.2	53.8
8	Demographics (Nominal)	7	Income, gender, profession	61.3	56.8

From Table 20, it can be seen that there are different variables extracted by CART in each model. To check if the best accuracy is achieved by those models, that is, to lower the possibility that more noise would be generated by many variables, two combined mode which only use the extracted variables by those 7 models are proposed. The mode I only apply the variables discovered by model 1, 2 and 3, and the mode II only apply the variables discovered by model 4, 5 and 6, the results of two combined mode are shown in Table 21 and Table 22. And the tree of the best trial of each mode is also shown in Fig. 16 and Fig. 17. From the combined mode, it can be seen that the variable extracted and the accuracy of classification did not have a significant change or growth; it means that the result of model 7 did not have a strong influenced by variable numbers. This proves that CART can still extract the effective variables through large number of variables.

Table 21 Combined mode I for digital camera

Trial	**Variable selection**	**Accuracy of classification (Training) (%)**	**Accuracy of classification (Testing) (%)**
1	**(V104, V12)**	61.4	52.2
2	(V64)	57.3	47.0

3	**(V104)***	58.3	58.2
4	**(V104)**	59.7	55.0
5	(V104, V64, V29)	64.0	48.2
6	**(V52)**	57.8	51.8
7	**(V64, V29)**	60.2	51.0
8	**(V8)**	56.8	51.2
9	**(V104)**	58.4	58.1
10	**(V29)**	58.9	52.9

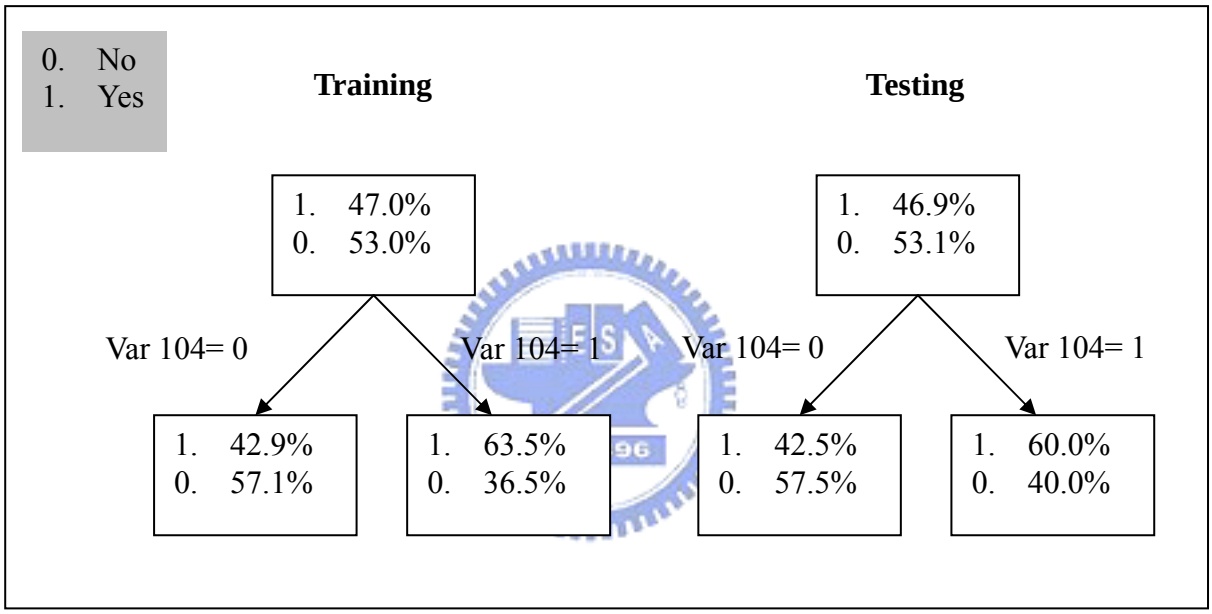


Fig. 16 The marketing segmentation of digital camera by CART-the best trial of combined mode I

Table 22 Combined mode II for digital camera

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	(V33)	59.0	49.5
2	**(V104)**	60.5	54.0
3	**(V104, V64)**	60.6	56.3
4	**(V104, V64)***	59.8	58.5
5	**(V104, V64)**	63.0	52.1
6	**(V104, V64)**	61.2	51.

7	**(V104, V64, V73)**	61.7	53.8
8	**(V104)**	60.4	53.6
9	**(V64, V73)**	60.7	58.0
10	(V63, V33)	61.6	41.3

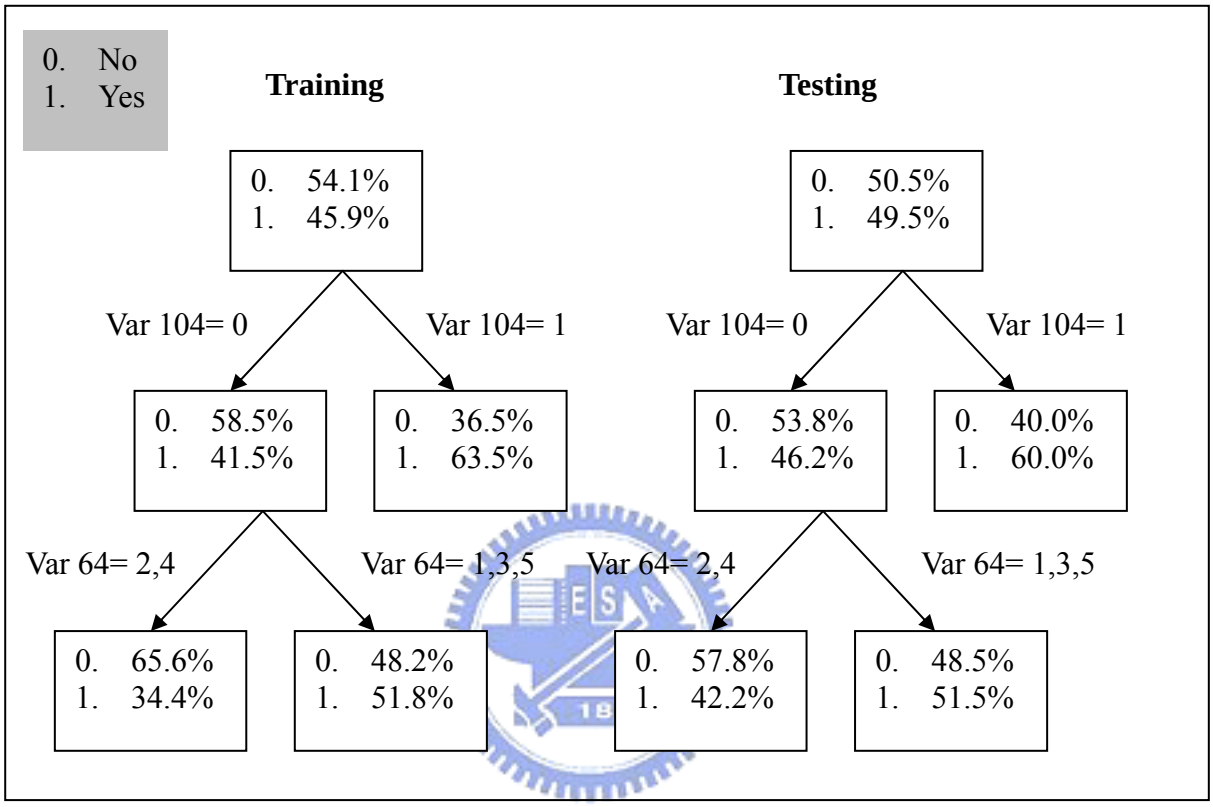


Fig. 17 The marketing segmentation of digital camera by CART-the best trial of combined mode II

4.3.2 Empirical results –Book online

In the model 1, 58 general lifestyle variables in Likert scale are the input variables. After computing process of CART, the results of 10 times trials are shown in Table 23. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V58), (V2, V58), (V2) and (V55) are effective explanatory variables which can used to predict the purchasing behavior of buying books online, the details of the result of those segmentation variables are shown in Table 24. Among those extracted variables, (V2) are the variable with the highest accuracy of testing and can be viewed as the best result of all trials, the

training and testing tree of this trial is shown in Fig. 18. From the tree, it can be explained that people who are not completely agree with the traditional thought that women's main mission is to provide happiness to her family members have more possibilities to buy books online.

Table 23 Experimental results of online book purchasing behavior using ordinal lifestyle variable by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V58)**	64.6	50.5
2	**(V58)**	62.2	56.8
3	**(V2, V58)**	64.0	53.3
4	**(V58)**	58.9	58.7
5	**(V2)***	62.1	60.6
6	**(V58)**	62.7	54.9
7	**(V58)**	61.8	57.7
8	**(V55)**	61.0	54.5
9	**(V2)**	63.7	56.4
10	**(V2)**	62.2	60.2

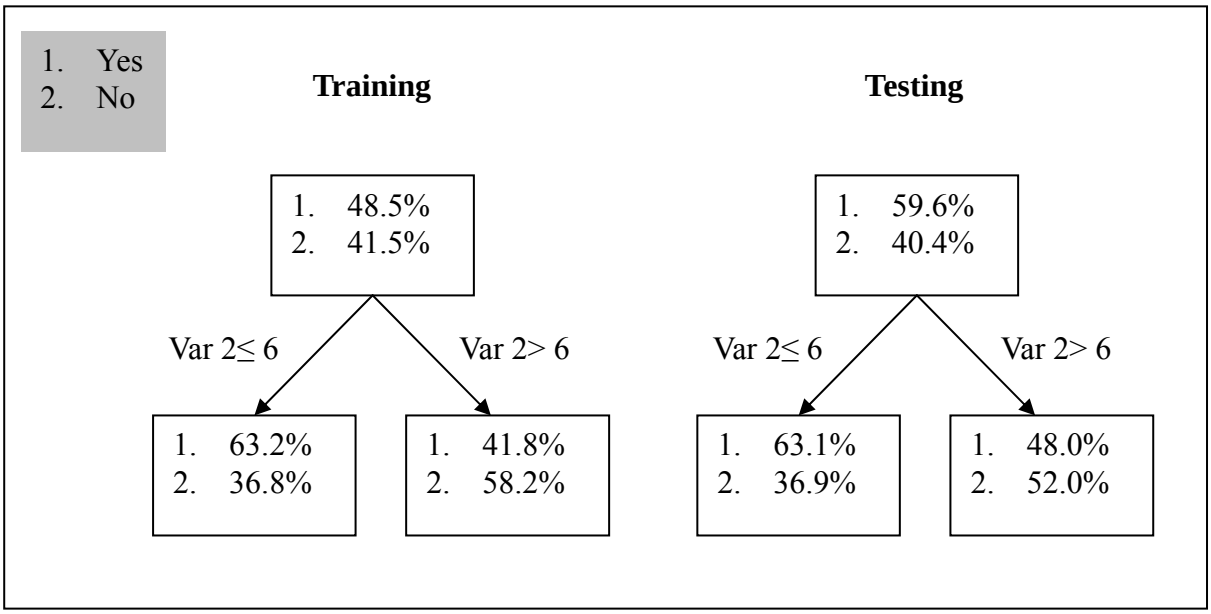


Fig. 18 The marketing segmentation of buying books online by CART-the best trial of model 1

Table 24 The result of market segmentation of purchasing books online by CART- model 1

Model 1	Variables extracted by CART and Classification results	Buying Books online
(V58)	V58: I have help many people to solve the doubt when using technological products	
	V58>4	Yes
	V58≤4	No
(V2, V58)	V2: the main task for woman is to provide happiness to her family members V58: I have help many people to solve the doubt when using technological products	
	V2>6	No
	V2≤6 + V58>4	Yes
	V2≤6 + V58≤4	No
(V2)	V2: the main task for woman is to provide happiness to her family members	
	V2>6	No
	V2≤6	Yes
(V55)	V55: I will make purchase decision after checking if there are small gifts or promotions of the products	
	V55≤4	Yes
	V55>4	Yes

In the model 2, 10 general lifestyle variables in nominal scale are the input variables. After computing process of CART, the results of 10 times trials are shown in Table 25. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V62), (V66, V62), (V62, V64) and (V66) are effective explanatory variables which can use to predict the purchasing behavior of buying books online, the details of the result of those segmentation variables are shown in Table 26. Among those extracted variables, (V62) are the variable with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 19. From the classification

tree, it can be explained that people who spent more time using internet, that is, more than 20 hours per week have more possibilities to buy books online.

Table 25 Experimental results of online book purchasing behavior using nominal lifestyle variable by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V62)**	60.2	55.3
2	**(V62)**	59.8	54.1
3	**(V62)***	58.9	58.7
4	**(V66, V62)**	59.9	56.4
5	**(V62, V64)**	63.3	55.0
6	**(V62)**	60.0	55.9
7	**(V62)**	59.3	57.7
8	**(V62)**	61.1	54.5
9	**(V62)**	61.0	54.5
10	**(V66)**	60.2	55.6

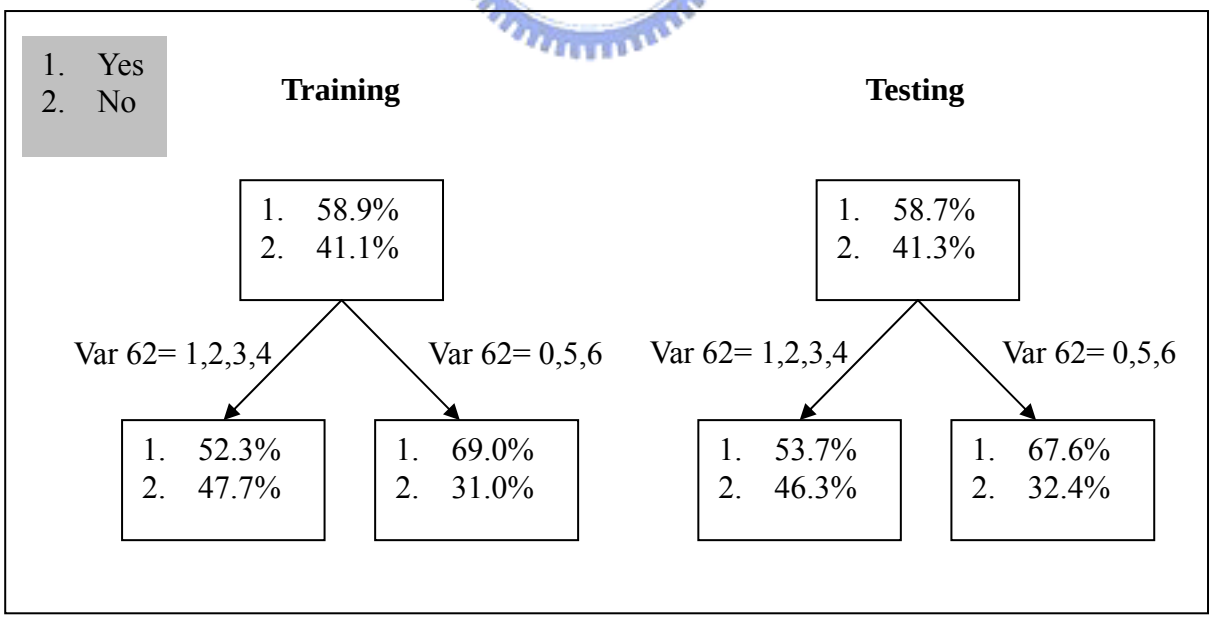


Fig. 19 The marketing segmentation of buying books online by CART-the best trial of model 2

Table 26 The result of market segmentation of purchasing books online by CART- model 2

Model 2	Variables extracted by CART and Classification results	Buying Books

		online
(V62)	V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours	
	V62=1,2,3	No
	V62=4,5,6,0	Yes
(V66, V62)	V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours V66: Are you willing to exchange your 10 years life to double your assets after retirement? 1: Yes; 2: No; 3: Can't decide	
	V66=1, 3	Yes
	V66=2 + V62=2,4	Yes
	V66=2 + V62=1,3,5,6	Yes
(V62, V64)	V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others	
	V62=5,6	Yes
	V62=1,2,3,4 + V64=2,5	Yes
	V62=1,2,3,4 + V64=1,3,4	No
(V66)	V66: Are you willing to exchange your 10 years life to double your assets after retirement? 1: Yes; 2: No; 3: Can't decide	
	V66=1, 3	Yes
	V66=2	Yes

In the model 3, 52 general lifestyle variables in nominal scale are the input variables.

After computing process of CART, the results of 10 times trials are shown in Table 27. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V94, V108), (V94), (V108), (V74, V80), (V94, V80), (V106, V74), (V114) and (V80) are effective explanatory variables which can use to predict the purchasing behavior of buying books online, the details of the result of those segmentation variables are shown in Table 28. Among those extracted variables, (V106, V74) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 20. From the tree, it can be explained that people who like to cook and don't like to go to KTV have more possibilities to buy books online.

Table 27 Experimental results of online book purchasing behavior using nominal activities variable by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V94, V108)**	63.6	52.4
2	**(V94)**	61.8	56.2
3	**(V108)**	59.2	58.5
4	**(V94)**	61.0	59.1
5	**(V74, V80)**	59.8	51.1
6	**(V94, V80)**	61.9	56.9
7	**(V106, V74)***	61.9	61.5
8	(V114, V80)	65.1	49.5
9	**(V114)**	60.8	54.7
10	**(V80)**	57.6	61.0

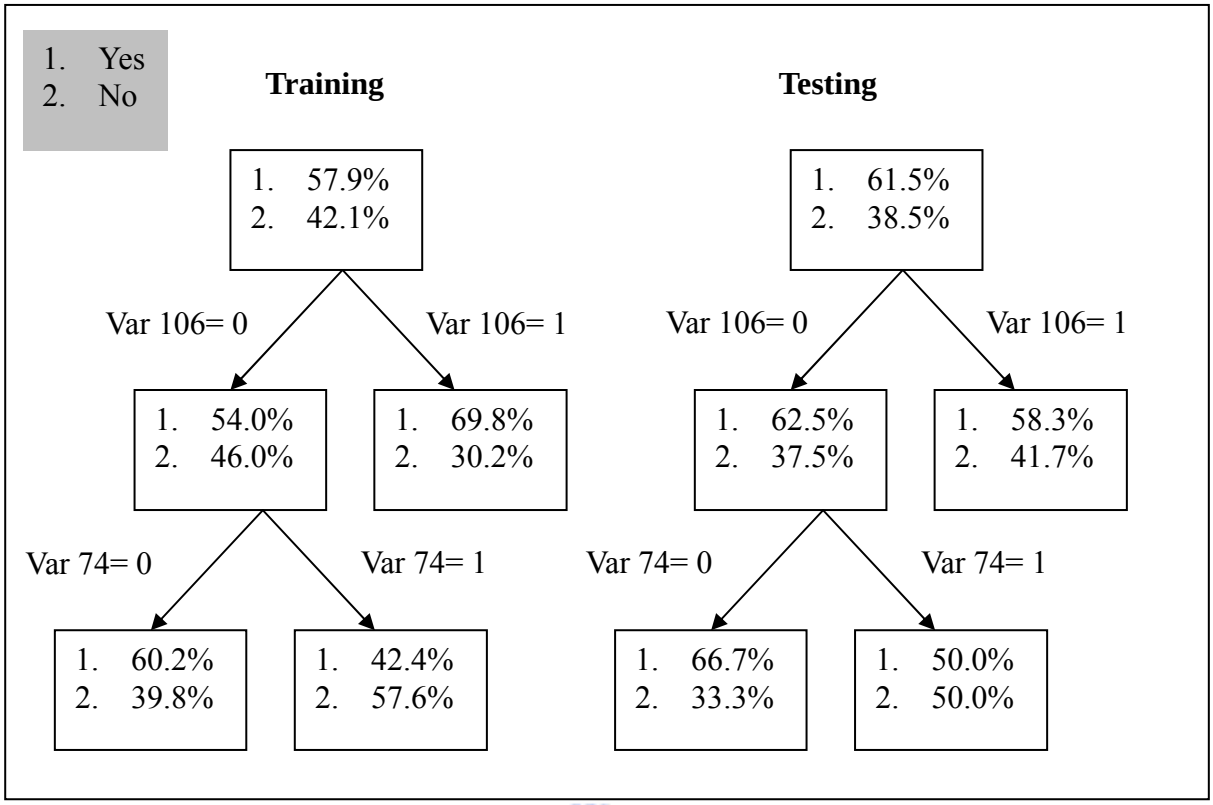


Fig. 20 The marketing segmentation of buying books online by CART-the best trial of model 3

Table 28 The result of market segmentation of purchasing books online by CART- model 3

Model 3	Variables extracted by CART and Classification results	Buying Books online
(V94, V108)	V94: Play basketball 0: No; 1: yes V108: Learn languages 0: No; 1: yes	
	V94=1 V94=0 + V108=0 V94=0 + V108=1	No Yes Yes
(V94)	V94: Play basketball 0: No; 1: yes	
	V94=0 V94=1	Yes No

(V108)	V108: Learn languages 0: No; 1: yes	
	V108=0 V108=1	Yes Yes
(V74, V80)	V74: Go to KTV 0: No; 1: yes V80: Eat in the restaurant 0: No; 1: yes	
	V74=1 V74=0 + V80=0 V74=0 + V80=1	No Yes Yes
(V94, V80)	V94: Play basketball 0: No; 1: yes V80: Eat in the restaurant 0: No; 1: yes	
	V94=1 V94=0 + V80=0 V94=0 + V80=1	No Yes Yes
(V106, V74)	V106: Cooking 0: No; 1: yes V74: Go to KTV 0: No; 1: yes	
	V106=1 V106=0 + V74=0 V106=0 + V74=1	Yes Yes No
(V114)	V114: Accompany the family 0: No; 1: yes	

	V114=0	Yes
	V114=1	Yes

In the model 4, 58 general lifestyle variables in Likert scale plus 10 general lifestyle variables in nominal scale are the input variables, after computing process of CART, the result of 10 times trials are shown in Table 29. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V14, V62), (V2), (V58) (V2, V56), (V62, V34) and (V4, V58) are effective explanatory variables which can use to predict the purchasing behavior of buying books online, the details of the result of those segmentation variables are shown in Table 30. Among those extracted variables, (V2) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 21. From the tree, it can be explained that p people who are not completely agree with the traditional thought that women's main mission is to provide happiness to her family members have more possibilities to buy books online.

Table 29 Experimental results of online book purchasing behavior using ordinal and nominal lifestyle variables by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	(V14)	61.7	49.1
2	(V34)	64.4	48.6
3	(V14)	61.5	49.1
4	**(V14, V62)**	64.4	61.0
5	**(V2)**	64.2	55.9
6	**(V58)**	62.6	56.1
7	**(V2, V56)**	63.2	57.6
8	**(V62, V34)**	60.4	52.3
9	**(V4, V58)**	63.1	52.1
10	**(V2)***	61.6	61.7

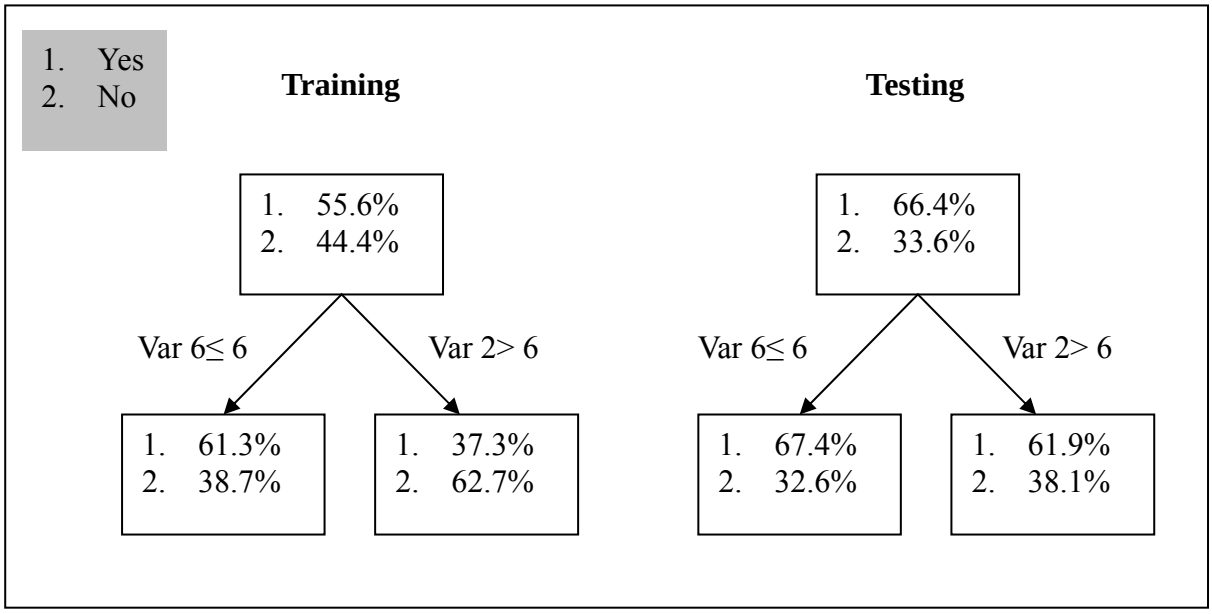


Fig. 21 The marketing segmentation of buying books online by CART-the best trial of model 4

Table 30 The result of market segmentation of purchasing books online by CART- model 4

Model 4	**Variables extracted by CART and Classification results**	**Buying Books online**
(V14, V62)	V14: I like to spend money to enjoy good food in the restaurant V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours	
	V14>6	Yes
	V14≤6 + V62=1,2,4	No
	V14≤6 + V62=3,5,6	Yes
(V2)	V2: the main task for woman is to provide happiness to her family members	
	V2>6	No
	V2≤6	Yes
(V58)	V58: I have help many people to solve the doubt when using technological products	
	V58>4	Yes
	V58≤4	No

(V2, V56)	V2: the main task for woman is to provide happiness to her family members V56: I am enthusiastic to cope with the affairs of neighborhood or community	
	V2>6	No
	V2≤6 + V56≤4	Yes
	V2≤6 + V56>4	Yes
(V62, V34)	V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours V34: I pay a lot of attention to my performance of my study or work	
	V62=1,2,3	No
	V62=4,5,6,0 + V34≤6	Yes
	V62=4,5,6,0 + V34>6	Yes
(V4, V58)	V4: I like to join the activities hold by the community V58: I have help many people to solve the doubt when using technological products	
	V4>4	No
	V4≤4 + V58≤4	Yes
	V4≤4 + V58>4	Yes

In the model 5, 10 general lifestyle variables (interest and opinion)in nominal scale plus 52 general lifestyle variables (activities) are the input variables, after computing process of CART, the result of 10 times trials are shown in Table 31. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V62), (V108, V62), (V94), (V94, V62), and (V94, V64) are effective explanatory variables which can use to predict the purchasing behavior of buying books online, the details of the result of those segmentation variables are shown in Table 32. Among those extracted variables, (V62) are the variable with the highest accuracy of testing and can be viewed as the best result of all trials, the

training and testing tree of this trial is shown in Fig. 22. From the tree, it can be explained that people who spent more time using internet, that is, more than 20 hours per week have more possibilities to buy books online.

Table 31 Experimental results of online book purchasing behavior using nominal lifestyle and nominal activities variables by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V62)**	60.2	55.3
2	**(V108, V62)**	65.0	50.5
3	**(V94)**	60.4	59.8
4	(V80)	60.1	46.2
5	**(V94, V62)**	63.2	53.6
6	**(V94, V64)**	61.3	57.8
7	**(V62)**	60.0	55.9
8	**(V62)**	61.1	54.5
9	**(V62)**	61.0	54.5
10	**(V62)***	58.2	60.4

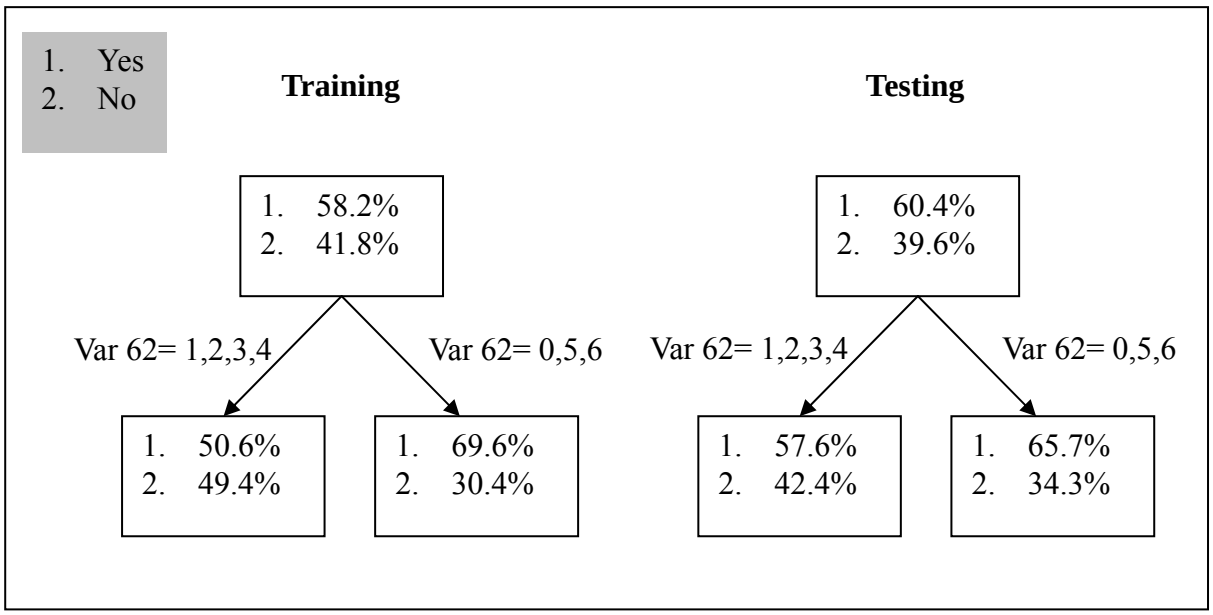


Fig. 22 The marketing segmentation of buying books online by CART-the best trial of model 5

Table 32 The result of market segmentation of purchasing books online by CART- model 5

Model 5	Variables extracted by CART and Classification results	Buying Books online
(V62)	V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours	
	V62=1,2,3	No
	V62=4,5,6,0	Yes
(V108, V62)	V108: Learn languages 0: No; 1: yes V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours	
	V108=1	Yes
	V108=0 + V62=1,2,3,5	No
	V108=0 + V62=0,4,6	Yes
(V94)	V94: Play basketball 0: No; 1: yes	
	V94=0	Yes
	V94=1	No
(V94, V62)	V94: Play basketball 0: No; 1: yes V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours	
	V94=1	No
	V94=0 + V62=1,2,3,4	Yes
	V94=0 + V62=5,6	Yes
(V94, V64)	V64: House types 1: House; 2: Apartment; 3: Mansion; 4: Suite; 5: Others	

	V94=1	No
	V94=0 + V64=1,3,4	Yes
	V94=0 + V64=2,5	Yes

In the model 6, 58 general lifestyle variables (interest and opinion) in Likert scale plus 52 general lifestyle variables (activities) are the input variables. After computing process of CART, the results of 10 times trials are shown in Table 33. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V2, V58), (V2), (V58), (V58, V80), (V58, V11), (V58, V34) and (V20) are the effective explanatory variables which can use to predict the purchasing behavior of buying books online, the details of the result of those segmentation variables are shown in Table 34. Among those extracted variables, (V58, V11) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 23. From the tree, it can be explained that people who seldom help other people solve the problem of using technological products have less possibilities to buy books online, on the other hand, people who help people solve the problem and don't really like social life will have more chance to by books online.

Table 33 Experimental results of online book purchasing behavior using nominal lifestyle and nominal activities variables by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V2, V58)**	64.0	53.3
2	**(V2)**	64.4	55.5
3	**(V58, V2)**	64.9	57.3
4	**(V58)**	61.3	58.8
5	**(V58, V80)**	66.8	53.3
6	**(V58, V11)***	60.7	60.0
7	(V58)	58.6	45.1
8	**(V58, V34)**	62.5	56.0

9	**(V2)**	62.8	58.7
10	**(V20)**	59.4	57.4

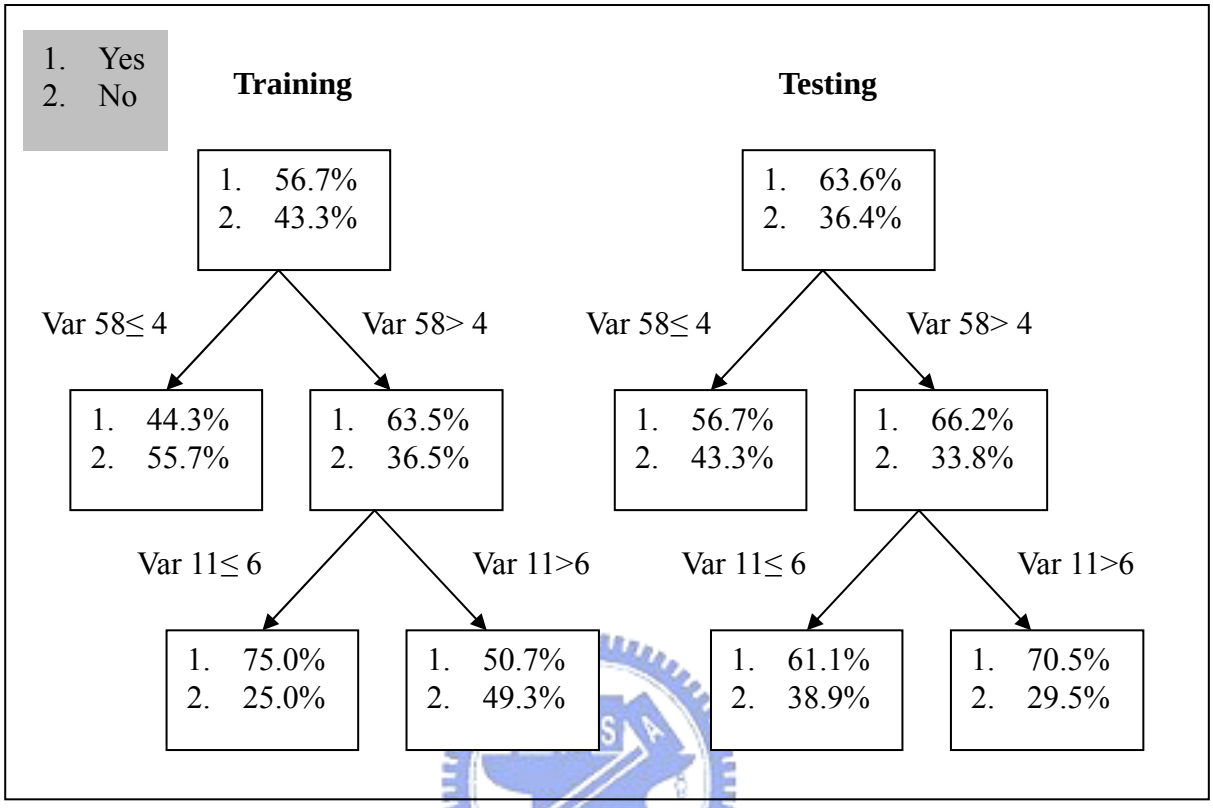


Fig. 23 The marketing segmentation of buying books online by CART-the best trial of model 6

Table 34 The result of market segmentation of purchasing books online by CART- model 6

Model 6	**Variables extracted by CART and Classification results**	**Buying Books online**
(V2)	V2: the main task for woman is to provide happiness to her family members	
	V2>6	No
	V2≤6	Yes
(V58, V2)	V58: I have help many people to solve the doubt when using technological products V2: the main task for woman is to provide happiness to her family members	
	V58>6	Yes
	V58≤6 + V2>6	No

	V58≤6 + V2≤6	Yes
(V58)	V58: I have help many people to solve the doubt when using technological products	
	V58>4	Yes
	V58≤4	No
(V58, V80)	V58: I have help many people to solve the doubt when using technological products V80: Eat in the restaurant 0: No; 1: yes	
	V58>4	Yes
	V58≤4 + V80=0	No
	V58≤4 + V80=1	Yes
(V58, V11)	V58: I have help many people to solve the doubt when using technological products V11: I like to know more friends and enjoy social life	
	V58≤4	No
	V58>4 + V11>6	Yes
	V58>4 + V11≤6	Yes
(V58, V34)	V58: I have help many people to solve the doubt when using technological products V34: I pay a lot of attention to my performance of my study or work	
	V58≤4	No
	V58>4 + V34>6	Yes
	V58>4 + V34≤6	Yes
(V20)	V20: I like to join politics activities	
	V20>2	Yes
	V20≤2	Yes

In the model 7, 58 general lifestyle variables (interest and opinion) in Likert scale, 10

general lifestyle variables (interest and opinion) in nominal scale and 52 general lifestyle variables (activities) are the input variables. After computing process of CART, the results of 10 times trials are shown in Table 35. In each trial, overall percentage of observation and prediction (training and testing) more than 50 percent will be selected, that is, the lifestyle variables extracted by those trials have explanatory power of purchasing behavior. Therefore, from the result of CART, (V94, V62), (V2) (V62), (V94), (V14, V62), (V58) and (V2, V56) are effective explanatory variables which can use to predict the purchasing behavior of buying books online, the details of the result of those segmentation variables are shown in Table 36. Between these two extracted variables, (V14, V62) are the variables with the highest accuracy of testing and can be viewed as the best result of all trials, the training and testing tree of this trial is shown in Fig. 24. From the tree, it can be explained that people who like to enjoy good food in the restaurants and spend more time on line per week have more possibilities to buy books online.

Table 35 Experimental results of online book purchasing behavior using ordinal and nominal lifestyle and nominal activities variables by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(V94, V62)**	63.7	51.9
2	**(V2)**	60.4	52.3
3	**(V62)**	60.7	52.2
4	(V14)	61.7	49.1
5	(V14)	61.5	49.1
6	**(V94)**	63.6	52.5
7	**(V14, V62)***	64.4	61.0
8	**(V2)**	64.2	55.9
9	**(V58)**	62.6	56.1
10	**(V2, V56)**	63.2	57.6

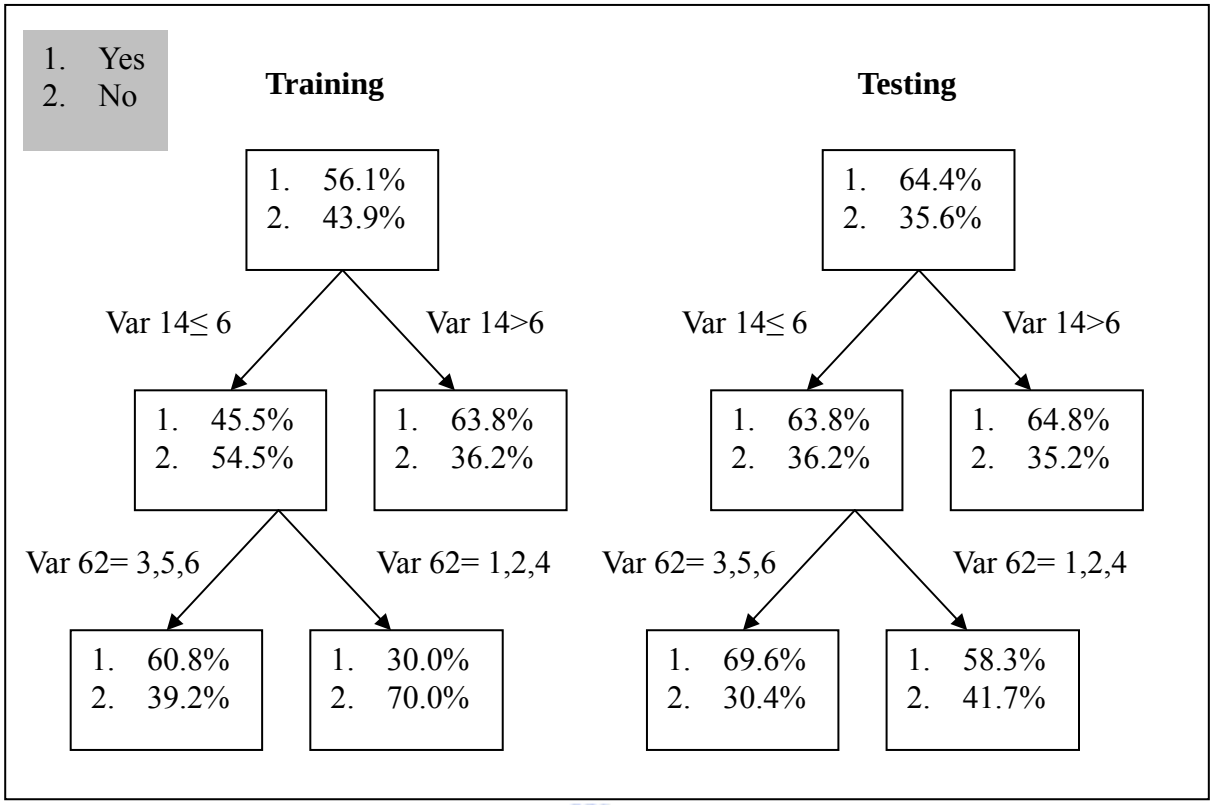


Fig. 24 The marketing segmentation of buying books online by CART-the best trial of model 7

Table 36 The result of market segmentation of purchasing books online by CART- model 7

Model 7	Variables extracted by CART and Classification results	Buying Books online
(V94, V62)	V94: Play basketball 0: No; 1: yes V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours	
	V94=1	No
	V94=0 + V62=1,2,3,4	Yes
	V94=0 + V62=5,6	Yes
(V2)	V2: the main task for woman is to provide happiness to her family members	
	V2>6	No
	V2≤6	Yes

(V62)	V62: Time spent online per week 1: Less than 5 hours; 2: 5~10 hours; 3: 11~15 hours; 4: 16~30 hours; 5: 21~25 hours; 6: more than 26 hours	
	V62=1,2,3	No
	V62=4,5,6,0	Yes
(V94)	V94: Play basketball 0: No; 1: yes	
	V94=0	Yes
	V94=1	No
(V14, V62)	V14: I like to spend money to enjoy good food in the restaurant V62: Time spent online per week 1. Less than 5 hours 2. 5~10 hours 3. 11~15 hours 4. 16~30 hours 5. 21~25 hours 6. more than 26 hours	
	V14>6	Yes
	V14≤6 + V62=1,2,4	No
	V14≤6 + V62=3,5,6	Yes
(V58)	V58: I have help many people to solve the doubt when using technological products	
	V58>4	Yes
	V58≤4	No
(V2, V56)	V2: the main task for woman is to provide happiness to her family members V56: I am enthusiastic to cope with the affairs of neighborhood or community	
	V2>6	No
	V2≤6 + V56≤4	Yes
	V2≤6 + V56>4	Yes

Table 37 Experimental results of purchasing books online using demographics variables by CART

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)
1	**(age, gender)**	63.2	52.3
2	**(age)**	60.7	58.2
3	**(age, gender)**	62.3	53.6
4	**(age, gender)**	62.0	55.8
5	**(age, gender)**	62.3	54.3
6	**(education, gender)**	62.7	57.8
7	**(education, gender)**	61.8	59.3
8	**(age, gender)**	60.2	59.4
9	**(age, gender)**	61.7	56.4
10	**(age, gender)**	62.5	53.8

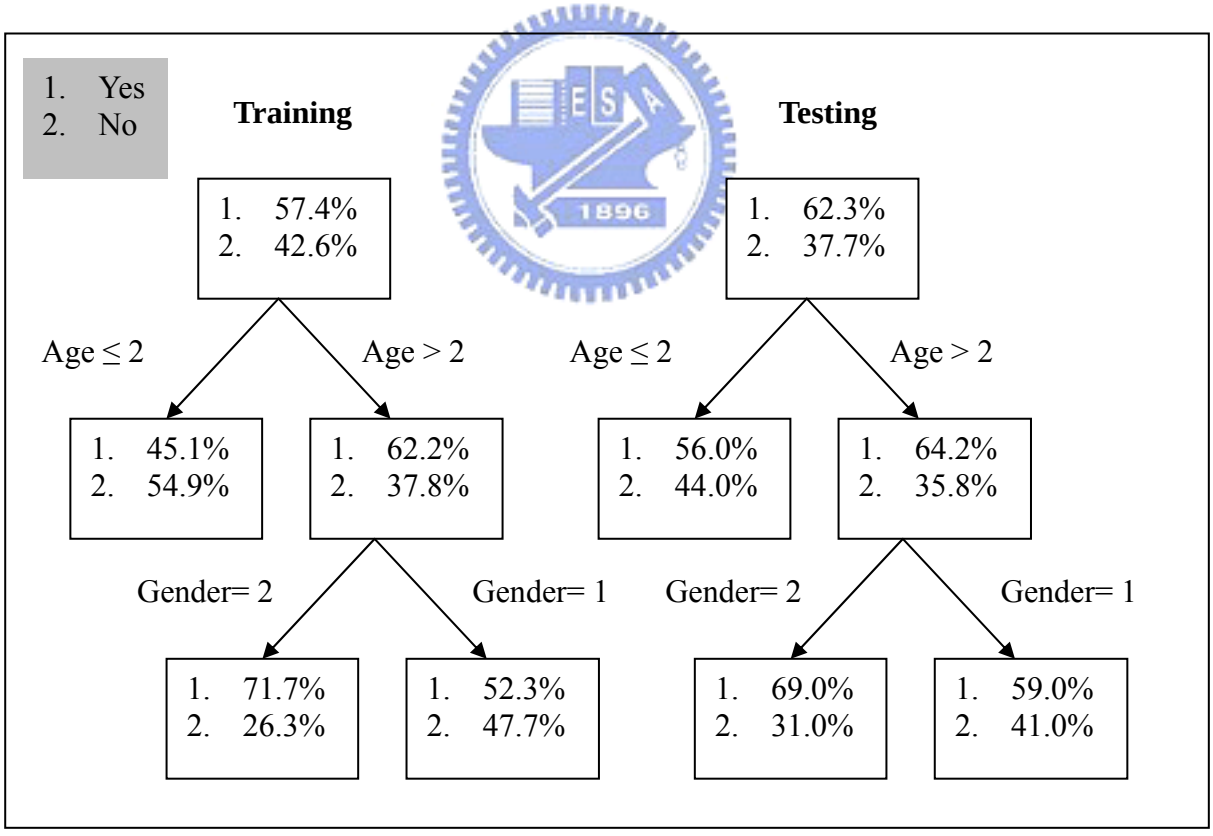


Fig. 25 The marketing segmentation of buying books online by CART-the best trial of model 8

Table 38 Summary of lifestyle variable selection of buying books online from different model by CART

model	Input variable	Number of variable	Total variable selected from the model	Accuracy of classification of the best trial (Training) (%)	Accuracy of classification of the best trial (Testing) (%)
1	Lifestyle(Scale)	58	V2, V58, V55	62.1%	60.6%
2	Lifestyle(Nominal)	10	V62, V66, V64	58.9%	58.7%
3	Activities(Nominal)	52	V94, V108, V74, V80, V106, V114	61.9%	61.5%
4	Lifestyle (Scale+ Nominal)	68	V14, V62, V2, V58, V56, V34, V4	61.6%	61.7%
5	Lifestyle(Nominal) + Activities (Nominal)	62	V62, V108, V94, V64	58.2%	60.4%
6	Lifestyle(Scale)+ Activities(Nominal)	110	V2, V58, V80, V11, V34, V20	60.7%	60.0%
7	Lifestyle(Scale, Nominal)+ Activities(Nominal)	120	V94, V62, V2, V14, V58, V56	64.4%	61.0%
8	Demographics (Nominal)	7	Age, gender, education	60.2%	59.4%

From Table 38, it can be seen that there are different variables extracted by CART in each model. To check if the best accuracy is achieved by those models, that is, to lower the possibility that more noise would be generated by many variables, two combined mode which only use the extracted variables by those 7 models are proposed. The mode I apply the variables discovered by model 1, 2 and 3, and the mode II apply the variables discovered by model 4, 5 and 6, the results of two combined mode are shown in Table 39 and Table 40. And the tree of the best trial of each mode is also shown in Fig. 26 and Fig. 27. From the combined mode, it can be seen that the variable extracted and the accuracy of classification did not have a significant change or growth; it means that the result of model 7 did not have a strong influenced by variable numbers. CART can still extract the effective variables through large number of variables.

Table 39 Combined mode I: Combined mode for buying books online

Trial	Variable selection	Accuracy of classification (Training)	Accuracy of classification (Testing)

		(%)	(%)
1	**(V62, V64)**	64.3	52.8
2	**(V58)**	61.9	58.7
3	**(V62)**	58.1	60.4
4	**(V2, V58)**	63.2	56.1
5	**(V62, V80)**	63.8	55.7
6	**(V62, V80)**	64.6	55.3
7	**(V94, V62)**	64.3	50.9
8	**(V2, V62)**	62.2	60.5
9	**(V62, V58)**	63.6	62.6
10	**(V62, V58)***	63.7	61.4

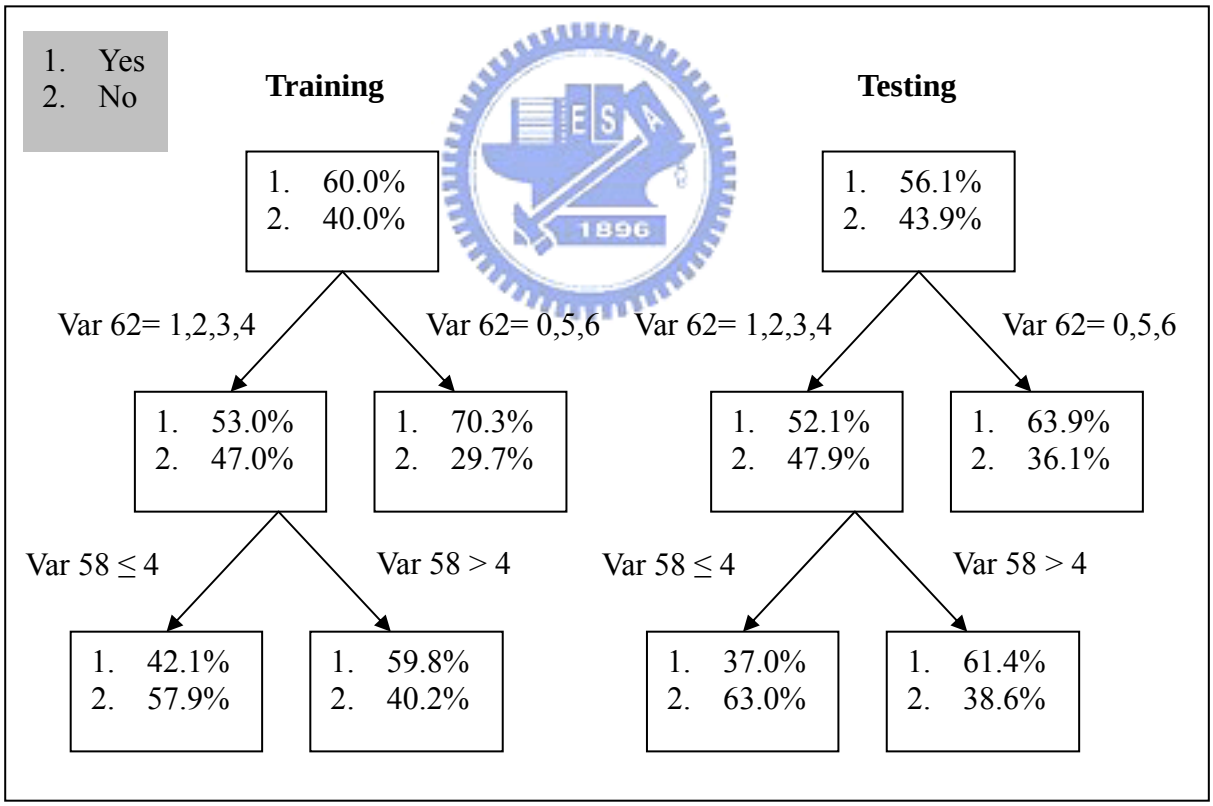


Fig. 26 The marketing segmentation of buying books online by CART-the best trial of combined mode I

Table 40 Combined mode II for buying books online

Trial	Variable selection	Accuracy of classification (Training) (%)	Accuracy of classification (Testing) (%)

1	**(V2)***	62.0	60.9
2	**(V62)**	60.1	53.2
3	**(V2, V56)**	59.6	50.5
4	**(V62)**	58.9	58.6
5	**(V2, V62)**	61.0	52.3
6	**(V58, V4)**	62.3	57.5
7	**(V2, V62)**	63.7	56.1
8	**(V94, V62)**	63.0	55.4
9	**(V58, V108)**	62.4	56.9
10	**(V94, V62)**	62.9	54.1

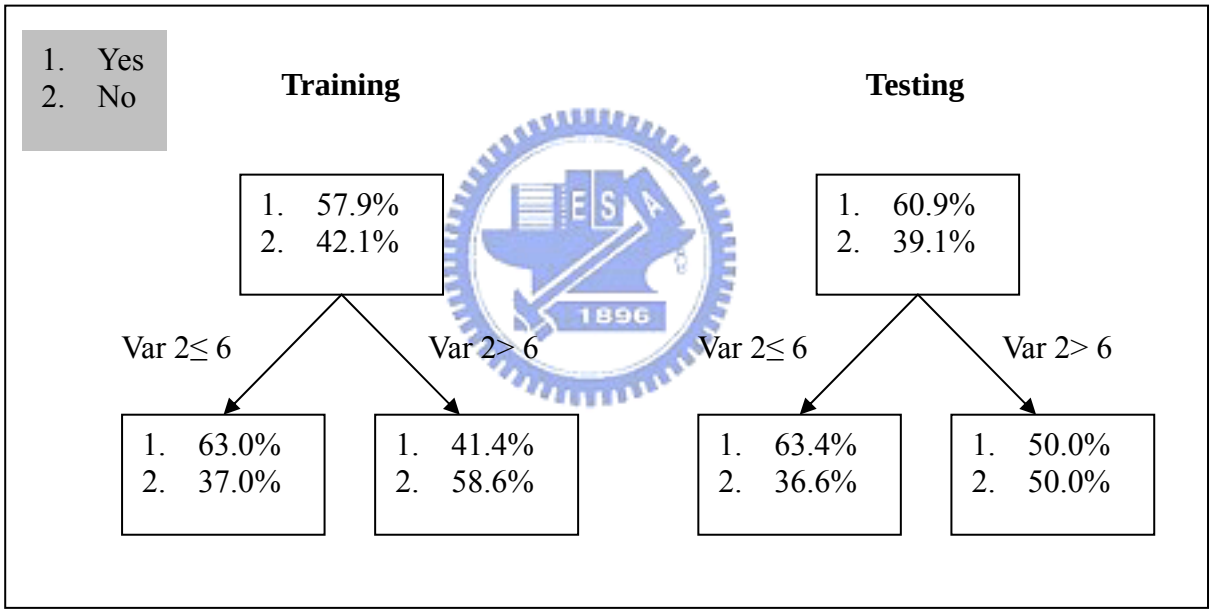


Fig. 27 The marketing segmentation of buying books online by CART-the best trial of combined mode II

4.3.3 The integrate model of CART and rough set

In this section, CART and rough sets are first used to derive the corresponding results of classification accuracy. Then, the results of CART and rough sets of used to compare with that of the proposed method.

To implement CART, processed data, i.e., remove noise and inconsistent data, are inputted to CART for the purpose of marketing segmentation, as shown in Fig. 28.

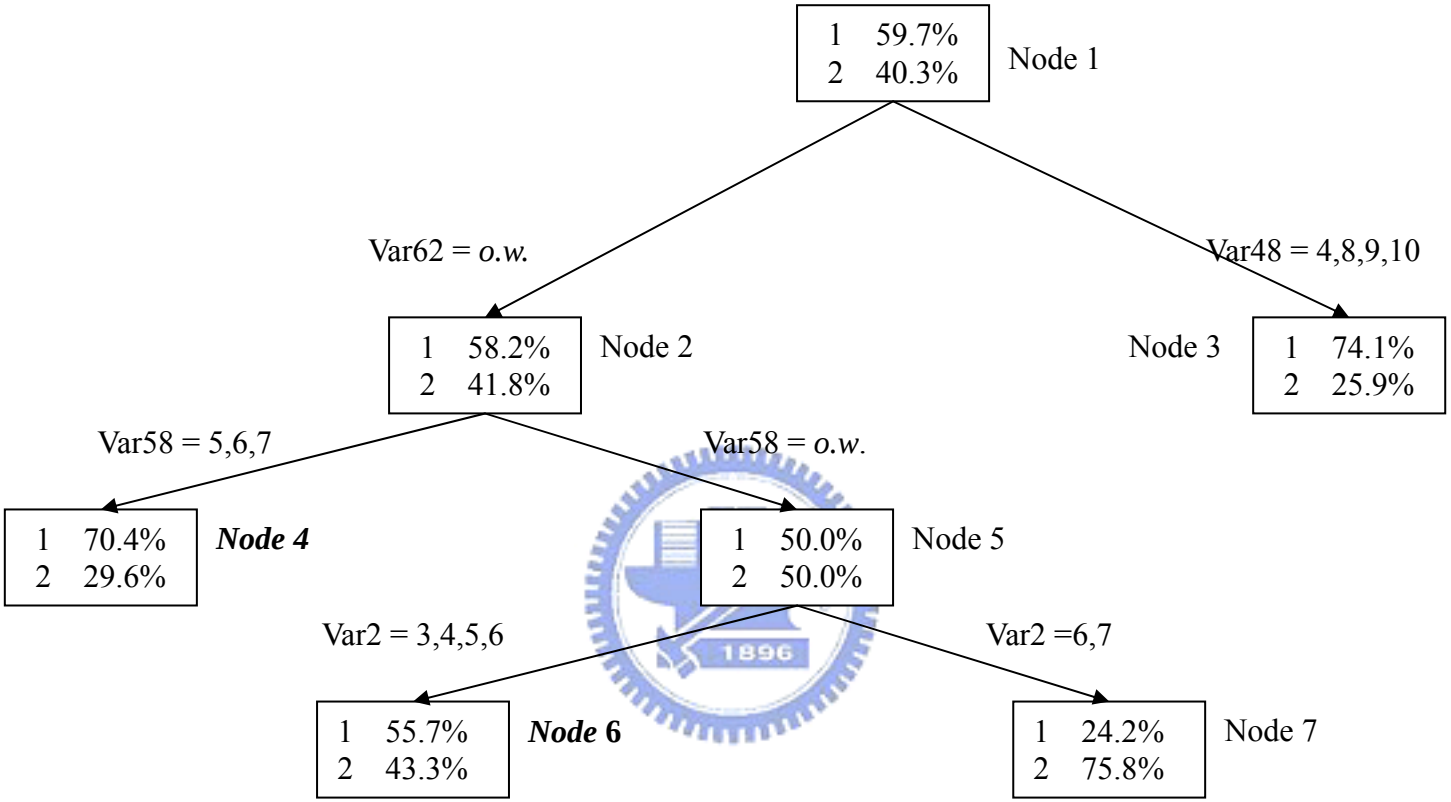


Fig. 28 The marketing segmentation of CART

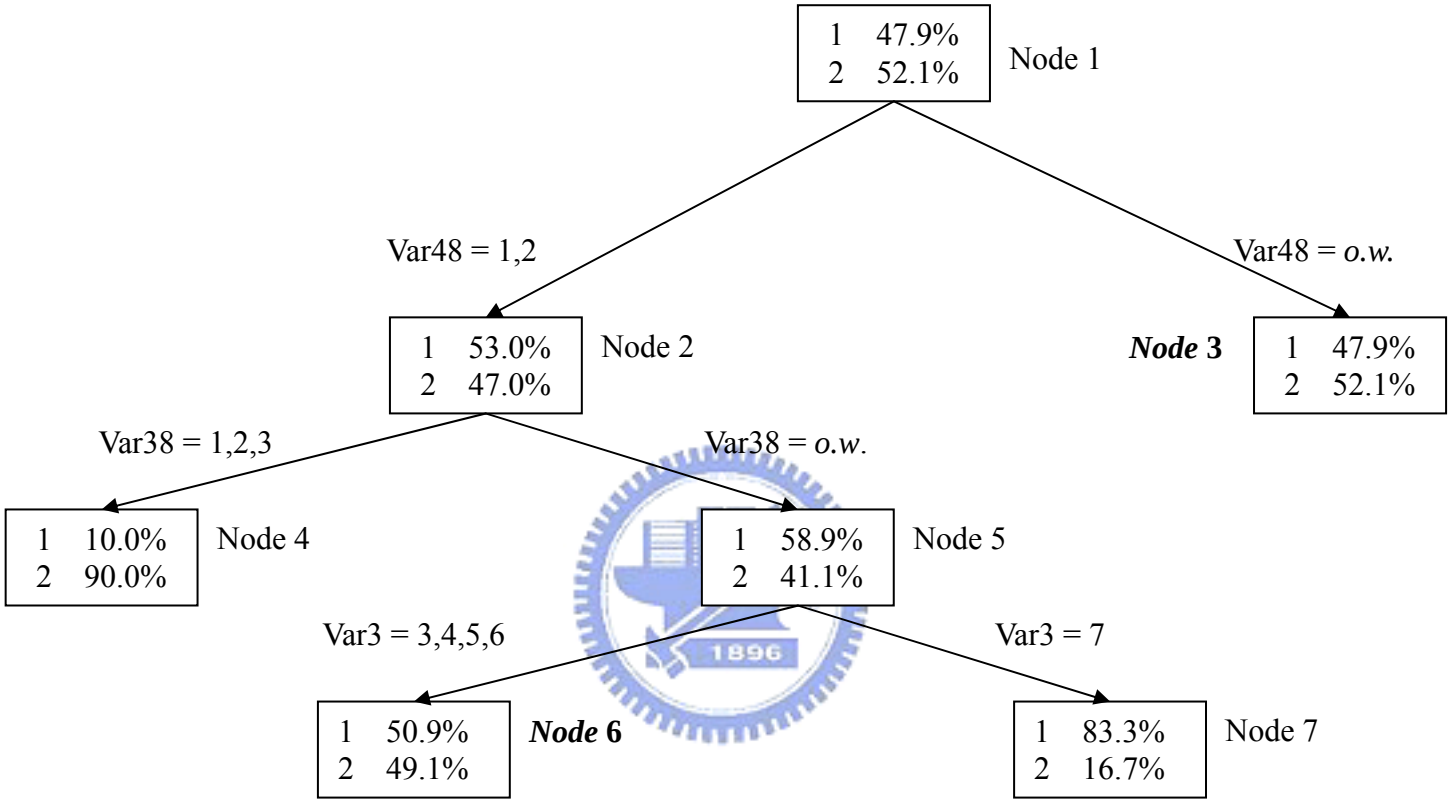


Fig. 29 The marketing segmentation of CART

Table 41The confusion matrix of CART.

		Prediction		
		1	2	Total
Actual	1	168	8	176
	2	94	25	119
	Total	262	33	295

On the other hand, if we input all processed into rough sets to classify whether a consumer may purchase a digital camera, we can obtain the classification accuracy of rough sets as 62.46%. The confusion matrix of rough sets can refer to Table 42.

Table 42 The confusion matrix of rough sets

		Prediction		
		1	2	Total
Actual	1	119	57	176
	2	80	39	119
	Total	199	96	295

From the result of CART, it can be seen that four segmentation of the digital camera, i.e., Nodes 3, 4, 6 and 7, can be identified. The corresponding classification accuracy of CART is equal to 56.63% and the confusion matrix of decision tree can be described as shown in Table 43.

Table 43 The confusion matrix of CART

		Prediction		
		1	2	Total
Actual	1	42	106	148
	2	28	133	161
	Total	70	239	309

On the other hand, if we input all processed into rough sets to classify whether a consumer may purchase a digital camera, we can obtain the classification accuracy of rough sets as 62.46%. The confusion matrix of rough sets can refer to Table 44.

Table 44 The confusion matrix of rough sets

		Prediction		
		1	2	Total
Actual	1	116	45	161
	2	71	77	148
	Total	187	122	309

From the results of Tables 43 and 44, it can be seen that the accuracy of rough sets is better than that of CART. However, rough sets are hard used for the purposed of marketing segmentation since the result of rough sets is composed of decision rules. Therefore, we pick the data of unsatisfied nodes, i.e., Nodes 3 and 6, to retrain via rough sets so that the accuracy of CART can be improved. However, it should be highlighted that number of marketing segmentation is reduced to three since Nodes 3 and 6 are integrated. Next, we can compare the accuracy of the proposed model with CART and rough sets as shown in Table 45.

Table 45 The comparison between CART, Rough sets and the proposed method

Comparison	CART	Rough Sets	The proposed method
Case 1	57.14%	53.56%	61.38%
Case 2	56.63%	62.46%	66.78%
Representation	Tree	Decision rules	Tree and decision rules
Fit	Continue data	Discrete data	Continue an discrete data
Mainly used for	Marketing segmentation	Classification	Both

From Table 45, we can conclude that the proposed method can appropriately

integrate the advantages of CART and rough sets with respect to the accuracy and marketing strategy. In other words, the proposed method can keep the characteristic of CART for the purpose of marketing segmentation and the accuracy can also be improved.

Chapter5 Conclusions

5.1 Research conclusions

This research aims to improve the traditional method used to segment the market by using lifestyle variables, on the one hand, lifestyle variables can enrich the information gathered of different market segments, on the other hand, using lifestyle variables as the basis of market segmentation, with the appropriate methodologies, different segments and consumer preferences can be discovered. From empirical analysis, it can be seen that this research have shown the explanatory power of general lifestyle variables to consumer behaviors. Even if there are few variables extracted by CART seems to be irrelevant, however, most of those lifestyle variables' relation to the specific behavior can be understood logically, furthermore, the segmentation results of demographic variables also provided in this research used to compare the explanatory power of general lifestyle variables. The empirical results show that, under different purchasing situation, the explanatory power of general lifestyle variables is not less than those traditional demographic variables; in fact, some models with different combinations of general lifestyle variables have higher explanatory power than demographic ones.

This research provide a novel model to improve and simplify the previous model proposed in lifestyle research, first, to reduce the excessive length of lifestyle questionnaires, those general lifestyle variables are measured with two different types of scale: ordinal and nominal, through this idea, the length of the questionnaire with general lifestyle can reduce from 200-300 items to about 100 items successfully. Second, to apply different types of scale in the questionnaire means higher complexity

of the analysis process, according to the literature review, we learn that there are many statistics tools need to be applied in this kind of situation, such as factor analysis, analysis of canonical correlations, cluster analysis and discrimination analysis, with the combination of all those methodologies, the model become very complicated and hard to interpret. Therefore, the appropriate methodologies CART and rough set which can deal with different types of variable scale are proposed in this research to develop the new model that integrates general lifestyle variables with different scales and appropriate methodologies, CART, can be consider the most important contribution of this research. This novel model not only can overcome the complexity happened to the previous research, in the mean time, it can be seen that CART actually perform decent ability to classify from the empirical results, therefore, we can conclude that this novel model can help the companies produce recommendation mechanism faster with accuracy.

CART can help to extract the variables with the best explanatory power with respect to specific purchasing behavior, in contrast with other previous general lifestyle researches, except the complexity of the model, using every general lifestyle variables to explain different behavior can be regarded as not necessary and redundant.

Even if the demographic variables have explanatory power to consumer behavior, however, it's not always easy to gather the real demographic information of the respondents; especially the companies want to sell products online, lifestyle questions are less related to privacy, after good packaging, these less aggressive questions let respondents have more willingness to provide the answers with better quality.

In this paper, a novel method, integrated by CART and rough sets, is propose to improve the shortcoming of CART. Although CART has been used for marketing segmentation, the result of CART is influenced by the complexity of data. Therefore,

in this paper, the concept of rough sets is incorporated to retain the data of unsatisfied nodes in CART to improve the classification accuracy of CART. From the comparison of CART, rough sets and the proposed method, we can conclude that the proposed method appropriately integrates the advantages of CART and rough sets. The integration of two methodologies, that is, CART and rough set, can increase the accuracy of classification effectively; also, it can improve the defects of CART and rough set when being applied to market segmentation.

5.2. Research limitation and suggestions

To prevent excessive variables to perplex the developed model in this research, we did not include the variables of values and personalities, however, according to previous researches, those variables are potentially with certain explanatory power of consumer behavior, therefore, these variables can also be added into the model in the future to help to improve the accuracy and richness of the market segmentation.

With the time and financial limitation to distribute the questionnaire results in uneven distribution of the sample, even though the sample collected is suitable for the objective behaviors of this research, for instances, college students are appropriate targets to explore the behavior of purchasing the digital camera and books online, even so, for various products, other groups of respondents could play critical roles to generate the accurate models, therefore, sample with fullness could create the model with higher discrimination power of buying behavior.

References

Ahmad, R. (2003). Benefit Segmentation: A potentially useful technique of segmenting and targeting older consumers. *International Journal of Market Research*, 45(3), 373-388.

Alpert, L. & Gatty, R. (1969). Product positioning by behavioral life-styles. *Journal of Marketing*, 33(2), 65-69.

Askegaard, S. & Brunsø, K. (1999). Food-related life styles in Singapore: Preliminary testingof a Western European research instrument in Southeats Asia. *Journal of Euromarketing*, 7(4), 65-86.

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

Barnes, G., Welte, J. & Dintcheff, B. (1991). Drinking among subgroups in the adult population of New York State: A classification analysis using CART. *Journal of Studies on Alcohol*, 52(4), 338-344.

Beatty, S. E., Kahle, L. R., Homer, P. & Mmisra, S. (1985). Alternative measurement approaches to consumer values: The list of values and the rokeach value survey. *Psychology and Marketing*, 2(3), 181-200.

Becherel, L. & Vellas, F. (1999). *The international marketing of travel and tourism-A strategic approach*. London: Macmillan Press.

Beynon, M. J. & Peel, M. J. (2001). Variable precision rough set theory and data discrimination: An application to corporate failure prediction. *OMEGA: the International Journal of Management Science*, 29(6), 561-576.

Boerstler, H. & de Figneiredo, J. M. (1991). Prediction of use of psychiatric services application of the CART algorithm. Journal of Mental Health Administration, 18(1), 27-34.

Boote, A. S. (1981). Market segmentation by personal values and salient product attributes. *Journal of Advertising Research*, 21(1), 29-35.

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

Cho, Y. H., Kim, J. K. & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Applications*, 23(3), 329-342.

Clark, L.A. & Pregibon, D. (1992). Tree-based models. In: Statistical Models in S, Chambers, J.M. Hastie, T.J., (Eds.), *Wadsworth Brooks/Cole, Pacific Grove, CA*.

Craig T. J., Siegel, C., Hopper, K., Lin, S. & Sartorius, N. (1997). Outcome in schizophrenia and related disorders compared between developing and developed countries: A recursive partitioning re-analysis of the WHO DOSMD data. British Journal of Psychiatry, 170(3), 229-233.

Darden, W. R. & Reynolds, F. D. (1971). Shopping orientations and product usage rates. *Journal of Marketing Research*, 8(4), 505–508.

Dibb, S., Simkin, L. & Bradley, J. (1996). The marketing planning workbook, *International Thomson Business Press, London*.

Dibb, S., Stern, P., Wensley, R. (2002), "Marketing knowledge and the value of segmentation", Marketing Intelligence and Planning, Vol. 20 No.2, pp.113-9.

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.

Dolnicar, S. B. (2004). Commonsense segmentation-a systematics of segmentation approaches in tourism. *Journal of Travel Research*, 42(3), 244–250.

Eric, H. S., Keith, L. M. & Cheep, K. O. (2000). The decision tree approach to stock selection. *Journal of Portfolio Management*, 27(1), 42-52.

Fisher, A. B. (1990). What consumers want in the 1990s. *Fortune*, 121(3), 108-112.

Fisher, R. J. (1990). *The social psychology of intergroup and international conflict resolution*. New York: Springer-Verlag.

Fournier, S., Antes, D. & Beaumier, G.. (1992). Nine consumption lifestyles. *Advances in Consumer Research*, 19(1), 329-337.

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

Gunter, B. & Furnham, A. (1992). *Consumer profiles: An introduction to psychographics*. Routledge: New York.

Grunert, K. G., Brunsø, K. & Bisp, S. (1997). Food-related life style: Development of a cross-culturally valid instrument for market surveillance. In: L. R. Kahle & L. Chiagouris (Eds.),Values, lifestyles and psychographics, *Mahwah, NJ: Lawrence Erlbaum*, 337-354.

Goh, C. & Law, R. (2003). Incorporating the rough sets theory into travel demand analysis. *Tourism Management*, 24(5), 511-517.

Goldman, L., Cook, E. F. & Brand D.A. (1988). A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *N. Engl. J. Med.* 318(13), 797–803.

Gonzalez, A. M. & Laurentino, B. (2002). The construct lifestyle in market segmentation: The behavior of tourist consumers. *European Journal of Marketing*, 36(1/2), 51-85.

Heath, R. P. (1995). Psychographics: Q'est-ce q c'est. *Marketing Tools*, 74(7), 74-79..

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

Hunt, S. & Arnett, D.B. (2004). Market segmentation strategy, competitive advantage, and public policy: grounding segmentation strategy in resource-advantage theory. *Australasian Marketing Journal*, 12 (1), 7-26.

Inuiguchi, M. (2004). Generalizations of rough sets: From crisp to fuzzy cases. Rough sets and current trends in computing by Shusaku Tsumoto, Roman Slowinski, Jan Komoroski, Jerzy W. Grzymala- Busse (Eds.), *Lecture Notes in Artificial Intelligence (LNAI)*, 3066(1), 26–37.

Jackson, A. G., Leclair, S. R., Ohmer, M. C., Ziarko, W. & Al-kamhwi, H. (1996). Rough sets applied to materials data. ACTA Material, 44(11), 4475-4484.

Johnson, C. J., K. L. Parker, D. C. Heard, & M. P. Gillingham. (2002). A multiscale behavioral approach to understanding the movements of wood-land caribou. Ecological Applications, 12(6), 1840–1860.

Plummer, J. T. (1974). The concept and application of life style segmentation. *Journal of Marketing*, 38(1), 33-37.

Kahle, L.R. (1983). *Social values and social change: Adaptation to life in America*, New York: Praeger.

Kahle, L. R., Beatty, S. E. & Holmer, P. (1986). Alternative measurement approaches to consumer values: the list of values (LOV) and values and life style (VALS). *Journal of Consumer Research*, 13(3), 405-409.

Kamakura, W. P. & Wedel, M. (1995). Life-style segmentation with tailored interviewing. *Journal of Marketing Research*, 32(3), 308-321.

Kamakura, W. A. & Novak, T. P. (1992). Value-system segmentation: exploring the meaning of LOV. Journal of Consumer Research, 19(1), 119-32.

Kao, D. L. & Robert, D. S. (1999). Equity style time, *Financial Analysts Journa*l. 55(1), 37-48.

Kim, J. W., Lee, B. H., Shaw, M. J., Chang, H. L. & Nelson, M. (2001). Application of decision-tree induction techniques to personalized advertisements on internet storefronts International. *Journal of Electronic Commerce*, 5(3), 45-62.

Koivumaa-Honkanen, H., Koskenvuo, M., Honkanen, R. J., Viinamaki, H., Heikkila, K. & Kaprio, J. (2004). Life dissatisfaction and subsequent work disability in an 11-year follow-up. *Psychological Medicine*, 34(2), 221-228.

Kolter, P. (1997). Marketing management: Analysis, planning, implementation and control. *Englewood Cliffs, NJ: Prentice-Hall*.

Kotler, N. & Kotler, P. (1998). Museum strategy and marketing: Designing missions, building audiences, generating revenue and resources. *Jossey Bass.*

Kotler, P. (1984). Marketing management: Analysis, planning and control. *Prentice-Hall, Englewood Cliffs, NJ.*

Lambin, J. J. (1995). Marketing estrategico, 3rd ed., *McGraw-Hill*, Madrid."

Lazer, W. (1964). Life style concepts and marketing. In: S.A. Greyser, Editor, Toward scientific marketing, *American Marketing Association*, Chicago, IL, 243–252.

Li, R. & Wang, Z. O. (2004). Mining classification rules using rough sets and neural networks. *European Journal of Operational Research*, 157(2), 439-448.

Mair, J., Smidt, J., Lechleitner, P., Dienstl, F. & Puschendorf, B. (1995).A decision tree for the early diagnosis of acute myocardial infarction in nontraumatic chest pain patients at hospital admission. *Chest*, 108(6), 1502-1509.

Shaw, M. J., Subramaniam, C., Tan, G. W. & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1), 127-137.

Mitchman, R. (1991). *Lifestyle market segmentation*. Praeger, New York, NY.

Moschis, G. P. (1976). Shopping orientations and consumer uses of information. *Journal of Retailing*, 52(2), 61-70.

Novak, T. P. & MacEvoy, B. (1990). On comparing alternative segmentation schemes: The list of values (LOV) and values and life styles (VALS). *The Journal of Consumer Research*, 17(1), 105-109.

Orth, U. R., McDaniel, M., Shellhammer, T. & Lopetcharat, K. (2004). Promoting brand benefits: the role of consumer psychographics and lifestyle. *The Journal of Consumer Marketing*, 21(2/3), 97.

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. Rough sets and current trends in computing by Shusaku Tsumoto, Roman Slowinski, Jan Komoroski, Jerzy W. Grzymala-Busse (Eds.), *Lecture Notes in Artificial Intelligence (LNAI)*, 3066(1), 1-7.

Peppers, D. & Rogers, M. (1993). *The one to one future: building relationships one customer at a time*. New York: Currency and Doubleday

Polkowski, L. (2004). Toward rough set foundations, mereological approach. Rough sets and current trends in computing by Shusaku Tsumoto, Roman Slowinski, Jan Komoroski, Jerzy W. Grzymala-Busse(Eds.), *Lecure Notes in Artificial Intelligence (LNAI)*, 3066(1), 8-25.

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

Reynolds, F. D. & Darden, W. R. (1972). Intermarket patronage: a psychographic study of consumer outshoppers, *Journal of Marketing*, 36 (4), 50–54.

Richins, M. L. & Verhage B. J. (1985). Cross-cultural differences in consumer attitudes and their implications for complaint management. *International Journal of Research in Marketing*, 2(3), 197-206.

Ritchie, J. R. & Goeldner, C. (1987). *Travel, tourism, and hospitality research: A handbook for managers and researchers* ,Wiley, New York.

Shaw, C. E., Al-Chalabi, A. & Leigh, N. (2001). Progress in the pathogenesis of amyotrophic lateral sclerosis. *Curr Neurol Neurosci Rep.*;1:69–76.

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.

Swarbrooke, J. & Horner, S. (1999). *Consumer behaviour in tourism*. Butterworth-Heinemann, England.

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

Thach, L. & Olsen, J. (2006). Market segment analysis to target young adult wine drinkers. *Agribusiness: An International Journal,* 22(3), Special Issue on Wine Marketing.

Thomssen, C., Oppelt, P., Janicke, F., Ulm, K., Harbeck, N., Hofler, H., Kuhn, W., Graeff, H. & Schmitt, M. (1988). Identification of low-risk node-negative breast

cancer patients by tumor biological factors PAI-1 and cathepsin L. *Anticancer Research*, 18(3C), 2173-2180.

Valette-Florence, P. (1994). Introduction à l'analyse des chaînages cognitifs. *Recherche et Applications en Marketing*, 9(1), 93-117.

Van Raaij, W. F. & Verhallen, T. M. M. (1994). Domain-specific market segmentation. *European Journal of Marketing*, 28(10), 49–66.

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

Wedel, M. & Kamakura, W. A. (2000). Market segmentation. conceptual and methodological foundations. *International Series in Quantitative Marketing*, second ed. Kluwer Academic Publishers, Boston.

Wells, W. D. & Tigert, D. J. (1971). Activities, interests and opinions. *Journal of Advertising Research*, 11(4), 27-35.

Wendell, R. S. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of marketing*, 21(1), 3-8.

Wind, Y. & Green, P. E. (1974). Some conceptual, measurement and analytical problems in life style research. *Lifestyle and Psychographics*, American Marketing Association, Chicago, IL, 99-126.

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.

Witt, S. F. & Moutinho, L. (1994). *Tourism marketing and management handbook* (2nd ed.), Prentice Hall, New York.

Yohannes, Y. & Hoddinott, J. (1999). *Classification and regression trees: An introduction*. Technical Guide #3, IFPRI.

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

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

Appendix

Rough set algorithm for decision-making

Rough set theory is a mathematical approach to managing vague and uncertain data or problems related to information systems, indiscernibility relations and classification, attribute dependence and approximation accuracy, reduct and core attribute sets, and decision rules. The remainder of this section discusses the above areas in detail.

Information systems

Given a questionnaire model QM (an information system),

$QM = (U, A, V, \rho)$

$U = \{x_1, x_2, ..., x_n\}$

$A = \{\ features / attributes\} = \{a_1, a_2, a_3\}$

U	a_1	a_2	a_3
x_1	1	1	0
x_2	2	2	1
x_3	2	3	2
⋮	⋮	⋮	⋮
x_n	2	3	2

$V_{a_1} = \{1,2\}, V_{a_2} = \{1,2,3\}, V_{a_3} = \{0,1,2\}$

$V = \bigcup_{a \in A} V_a$

The information function, ρ, where U is the universal object sets of QM; A represents the model attribute sets, consisting of attributes $\{a_1, a_2, a_3\}$; V_{a_1} represents the domain (value sets) of attribute a_1; $V\left(=\cup_{a \in A} V_a\right)$ is a set of values of the attributes; $Ds(x) = \{f(x, a_1), f(x, a_2), ..., f(x, a_k)\}$ is the description of each object, x, of

U (Greco et al., 2001), and $f(x,a) \in V_a$ is called the information set of object x.

Therefore, $\cup Ds(x_1, x_2, x_3, \cdots, x_n)$ denotes the information of the system. Let $\rho : U \times A \to V$ be a description function, such that $\rho(x,a) \in V_a$ for each $a \in A$ and $x \in U$, where ρ_x is the description of x in QM (Pawlak, 2002).

If a single-choice question has a value set $V_{a_1} = \{1,2\}$ of attribute a_1, then the numeric value will be 2. if a multi-choice question has a value set $V_{a_2} = \{(1),(1,2),(1,3),(1,2,3),(2),(2,3),(3)\}$ of attribute a_2, then the numeric value will be $\sum_{i=1}^{n} C_i = 7$, where n is 3. The attribute a_2 is called the attribute of the combination value. The numeric value of attribute a_2 is double that of attribute a_1. Inducing the number of elementary sets of attributes of combination values doubles the attribute of a single value. This definitely derives a large discriminate data set for attributes of combination values, which makes classification more difficult. Each x in U may correspond with different or the same decision data (d_1, d_2, d_3). This induces the indiscernibility relation, which we discuss next.

Hereafter, we call the above table the decision table, and attributes are divided into condition attributes (denoted as CA) and decision attributes (denoted as DA).

Indiscernibility relation and classification

Let objects $x_1, x_2 \in U$ be indiscernible by the set of attributes B in A. 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 $\rho_{x_1}(a)$ for every $a \in A$. The equivalence class of $IND(B)$ is called an elementary set (atoms) in QM. Thus,

any x_i of U can be induced so that the value sets of attributes represented in B are in the same class. Objects grouped in the same class are called elementary sets, and the process is called classification. 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.

Attribute dependence and approximation accuracy

Let $QM = (U, A, V, \rho)$ be an information system, and let $a_1, a_2 \in A$. According to Pawlak (1984), 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*. Therefore, we can induce $IND(A) = IND(A - a_2)$, and a_2 is a superfluous attribute. The superfluous attributes are removed, which simplifies the information set and generates diagnostic values. In order 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. If the number of the elementary set is the same as the original set, the attribute is defined as a superfluous, and the remaining attributes are considered indispensable.

In decision rule extraction, the computation of accurate approximations is important. The intersection of conditions and decision classes yields both the lower and upper approximations. The expression is described as follows:

If X is U's subset, expresses objects $x_1, x_2, ..., x_n$ where i is 1 to n.

$$L_{app}(x_i) = \{x_i \in U \mid x_i \subseteq X\} \quad (1)$$

$$U_{app}(x_i) = \{x_i \in U \,|\, x_i \cap X \neq \phi\} \quad (2)$$

$$Bnd(x_i) = U_{app}(x_i) - L_{app}(x_i) \quad (3)$$

Eq. (1) represents the lower approximation and object x_i belongs to the elementary sets contained in X .Object x_i may, or may not, belong to the elementary sets contained in X that have non-empty intersections. This is called the upper approximation, as showen in Eq. (2). The difference between Eqs.(1) and (2) is expressed as Eq. (3), which is called the boundary region of X , indicating that the objects are inconsistent or vague. To sum up, the objects of $L_{app}(X_i) \subseteq$ objects of $U_{app}(x_i)$.

Reduct and core attribute sets

Reduct and core attribute sets are two fundamental concepts of rough set theory. Reducts are the most precise way of discerning object classes, which are the minimal subsets provided that the object classification is the sameas with the full set of attributes. The core is common to all reducts.

The reducts processor attributes reduces elementary set numbers, the goal of which is to improve the precision of decisions. After the attribute dependence process, the reduct attribute sets are generated to remove superfluous attributes, so that the set of attributes is dependent. 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 intersection a number of reduct attribute sets yields a core attribute set. The reduta 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.

$$RED(B) \subseteq A \tag{4}$$

$$COR(C) = \bigcap RED(B) \tag{5}$$

Based on the approximation method, the reduct attribute sets and decision rules can be derived such that the reduct set is a minimak set of attributes. Eq. (4) shows that A is the attribute set of U, and B is the reduct attribute set, so that B is included in, or equal to, A. In Eq. (5), the intersection of all reduct attribute sets is the core attribute set. Applying the reduct set to the model, we can induce the decision rules, which are based on the approximation method. The approximation accuracy rate is derived from the computation of the intersection rate between the lower and upper approximations, which are used to evaluate the classification's accuracy.

Decision rules

Let the attributes of set A be divided into CA and DA, where $CA \neq \phi$ and $DA \neq \phi$; then, $CA \cup DA = A$ and $CA \cap DA = \phi$, which are the elements of the decision table. DA induces an indiscernibility relation, $IND(DA)$, which is independent of CA. 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 ... $f(x, a_k)$, then x belongs to DA_1 or DA_2 or DA_n, where $\{a_1, a_2, ..., a_k\} \subseteq CA$ and $DA_1, DA_2, ..., DA_n$ are decision classes. If , then the rule is exact; otherwise, it is approximate or ambiguous (Greco et al., 2001).

$$\sigma_{QM}(\Phi,\Psi) = \sup p_{QM}(\Phi,\Psi) \Big/ card(U) \tag{6}$$

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. Eq. (6) is the strength of the decision rule $\Phi \to \Psi$ in QM, where the $\sup p_{QM}(\Phi,\Psi)$ is called the support of the rule $\Phi \to \Psi$ in QM, and $card(U)$ is the cardinality of a set, which is the number of objects contained in 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, i.e., they have shorter condition sets and fewer explanations. Dimitras et al. (1999) also propose 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.

Rough set theory usually assumes that there is only one decision attribute. If there is only one decision attribute. If there is more than one decision attribute, then different decision tables will be generated by the relation among decision attributes. 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. The type of decision table expressed in the terms of attributes CA can be expressed in the terms of attribute DA. Let an attribute set of $CA = \{a_1, a_2, a_3\}$ and an attribute set of $DA = \{d_1, d_2, d_3\}$, where decision attributes d_1 and d_2 are dependent on each other; then d_1, d_2 will form a set to generate a decision table, and the third attribute, d_3, will generate a table by itself.

This study focuses on the problem of classifying data sets into classes. 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. 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 data with attributes of combination values.

Source: 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.