Genetic algorithm based fuzzy association mining rules for means-end chain data

Nai-hua Chen*,
Assistant Professor,
Dept. of Information Management, Cheng-Kou University,
Chung-Hua, Taiwan. Email: nhc@ctu.edu.tw

Chi-Hsun Lee,
Professor,
Dept. of Business Management, National United University,
1 Lienda, Kung-ching Li, Maoli, 360, Taiwan. Email: cslee@nuu.edu.tw

Abstract

This paper aims to describe what attributes, consequences and value influence Taiwanese consumer‟s perception and adoption of organic rice through Means-end chain (MEC) model in a quantitative fashion, by introducing a new method of genetic algorithm based fuzzy association mining rules. With a sample of 250 green consumers, this study finds that Taiwanese consumers concern about nutrition, genuineness of food taste, pesticide-free, and food safety, by paying more attentions to production date, packaging, prize winning and the channel stores than other attributes. Beyond these functional consequences and self interests, the green consumers also emphasize harmony with others, the animal welfare, giving supports to organic farmers, along with the global trend which green consumption is fashionable. Echoing the List of Value (LOV) proposition, Taiwanese consumers search for the ultimate value of sense of fulfillment, accomplishment, security and fun and enjoyment of life from organic consumption. Furthermore, Taiwanese consumers

* Corresponding author, please contact her at nhc@ctu.edu.tw.

view healthier family as their ultimate value as well, which reflects the core values of eastern Asian cultural heritage.

1. Introduction

Consumers are not only buying things, but they also having and being themselves (Soloman, 2009), to express their extended self (Belk, 1988). What consumers consume reflect how they perceive the relationship between themselves and outer world around them, and what they concern, particularly true for food consumption, which is acquired by consumers through socialization process. As such, the conversion of food consumption is very important for consumers, since they abandon what they used to have as meals and change the habit or even ritual of food consumption to other ones. This paper aims to examine what influence Taiwanese consumers to convert their traditional food consumption to organic foods, and what the values beyond the consumption.

In marketing, values are often suggested as such broad motivational goals in the customer decision-making process. Consumers‟ product knowledge often derives from their cognition of values imbedded in product attributes or/and the consequences of using the product (Reynolds and Whitlark, 1995). Therefore, associations with product attributes and benefits can enable us to better understand customer product knowledge and purchasing motives. Such an approach, termed means-end chains, can be used for different purposes like improved segmentation (Botschen et al., 1999), or development of advertising strategy (Peter & Olson, 1999, Reynolds & Craddock, 1998). A means-end map arranges means and ends into a network of attributes, physical and emotional consequences, and personal values or life goals. The means-end framework for viewing consumer decision making leads to a means-end theory of communications strategy. Although theoretically clear, research of means-end structures is not an easy task. It requires the identification of various chain elements which encompass objective product entities (attributes), but also highly subjective constructs such as psychological benefits and values.

Association rule have been widely used to determine customer buying patterns from the market basket data. The mining association rule is to discover mainly association rules from large scale databases. The associate pattern technique is inspired by the Apriori mining algorithm in this study (Agrawal and Srikant, 1994). The Apriori algorithm is a classical algorithm for learning associate rule in the computer science field.

Fuzzy set theory has been used more and more frequently recently in intelligent system because of its simplicity and similarly to human reasoning. Many fields have applied the fuzzy set successfully, such as engineering, manufacturing and economics. Some fuzzy mining association algorithms are proposed in data mining. Hong et al.

introduce fuzzy mining algorithm for quantitative data (Hong et. al, 1999). Yue et al. use AprioriTis based algorithm to find association rules of quantitative transactions. Due to the time complexity in mining fuzzy association rule, this paper adopts the genetic algorithm to mine association rules. The purpose of this paper is to apply genetic algorithm based fuzzy association rule in a means-end approach to the organic product market. The aim of this paper is to provide suggestions for improvements in segmentation and advertising strategies.

2. Methodology

2.1 Means end chain

The „means-end chain‟ (MEC) framework ties these concepts together which could be represented in a hierarchical value map of purchasing motivation in Fig. 1 below. The hierarchy consists of six levels, from the most tangible to the most abstract: tangible product attributes, intangible product attributes, functional consequences, social psychological consequences from consuming the product, instrumental values, and terminal values from consuming the product. MEC offers a theoretical structure capable of linking consumers‟ values to product attributes and hence their buying behavior.

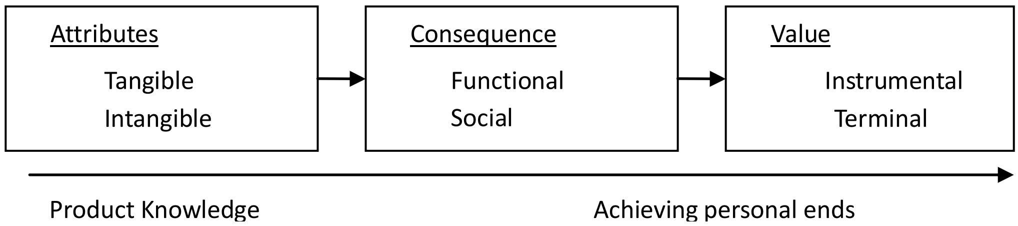

Figure. 1. Six levels of Means-end framework, adapted from Gutman, 1982.

Consumers‟ product knowledge often derives from their cognition of values imbedded in product attributes or/and the consequences of using the product (Reynolds and Whitlark, 1995). Consumer value formation may stem from the bundle of attributes that the product possesses which could be tangible or intangible (McColl-Kennedy and Kiel, 2000; Kotler, 2002). Product attributes included packaging, color, price, quality, brand, even the service level and reputation of the seller (Stanton et al., 1991; Pitts, et al., 1991). Product attributes are intermediate values that consumers perceive to achieve final ends of benefits rather than risks (Peter and Olson, 1993; Pieters, et al., 1995; Woodruff, 1997). In other words, many consumers tend to evaluate products based on the benefits arising from consuming the goods rather than based on product attributes alone.

In general, there are two kinds of consequences from using a product, functional and social psychological consequences (Gutman, 1982; McColl-Kennedy and Kiel, 2000). Functional consequences allow users to feel the direct and tangible benefits. Consumers‟ social psychological benefits from consuming the product are linked to how they personally perceive the value of the product. A person‟s general beliefs and values affect his/her attitudes toward specific things and often serves as a information filter (Beatty et al., 1991). Product values can be instrumental or terminal. Instrumental values are cognition of preferences or choices, such as versatility of life style, belief in independence, and confidence about the value that products or service may bring to the consumer. Terminal values are what the final state of products and services can bring to consumers. Examples of final state from consuming the product could be, peace of mind, emotional satisfaction and the like (Rokeach, 1968; Rokeach, 1973; Zanoli and Napetti, 2002). Many consumer studies have tried to measure personal values, including Rokeach Value Survey (RVS, Rokeach, 1973), Values and Lifestyles (VALS, Michell, 1983), List Of Values (LOV, Kahle et al., 1986).

2.2 The Association Mining Rules - Market Basket Analysis

The association mining rule is also called market basket analysis because it is useful to generate models for supermarket purchasing behavior and has been applied in consumer behavior analysis (Giudici and Passerone, 2002). It is an efficient data mining tool and represents a powerful method compared to log-linear method (Giudici, 2003). Association mining rules aim at extracting relevant knowledge from large database. The knowledge is usually expressed in the form of patterns of co-occurrence.

The most popular method to bring out co-occurrence of items in a database of transactions is the association rule.

The association rule is used in application such as market baskets analysis to measure the associations between products and customers. The association pattern technique used in this study is the Apriori mining algorithm. The steps of the Apriori algorithm is first finding the frequent itemsets from the transactional database. Second, removing low frequency itemsets based on the predefined minimum support level. Finally, the association rules are generated that satisfy the predefined minimum confidence from a given database (Agrawal and Srikant, 1994).

Two commonly used association rules are the support and the confidence. Normally in symbols, the expression of the association rule if A happens then B happens is written as $A \rightarrow B$, where A and B are attributes or factors under study. The definition of support is a relative frequency to indicate the proportion of both A and B have happened, or Supp $(A \rightarrow B) = |D_A \cap D_B| / |D|$, where $|D_A \cap D_B|$ is the transaction contains both A and B, and $|D|$ represents the cardinality of a database. The confidence rule is a relative frequency to indicate the proportion of transactions in which the support rule is observed given A has happened. The confidence rule is defined as Conf $(A \rightarrow B) = |D_A \cap D_B| / |D_A|$. The confidence indicates the conditional probability of B with respect to A.

2.3 Expanding the Association Mining Rule by Fuzzy Logic

The fuzzy set theory is widely used in intelligent systems for its ability to model inherent vagueness in nature. The framework of today"s fuzzy systems was developed by Zadeh in 1965 (Zadeh, 1965). Since then, scientists adopted the fuzzy system into applications systems in automatic control systems such as washing machines (Mendel, 1995). The generalization from classical to fuzzy association rules provides the ability to represents uncertainty and synergistic (Dubois, et al.).

The membership function may be s-type, z-type, triangular, trapezoidal, Gaussian etc. The membership function adopted in this study is the s-type membership function as we are interested in classifying whether respondents belong to the „important" or „unimportant" membership of 5-point Likert scale. The membership function is defined as

$$(x) = \begin{cases} 2(\frac{x-L}{H-L})^2, \text{if } L \leq x \leq \frac{L+H}{2} \\ 1 - 2(\frac{U-x}{H-L})^2, \text{if } \frac{L+H}{2} \leq x \leq H \end{cases}, \qquad (1)$$

where L is the lower bound and H is the upper bound.

The traditional Apriori algorithm explained above handles binary data, e.g., being belonging or not belonging to a set of membership under observation. However, in real life application databases may contain other attribute values besides 0 and 1, (e.g., 1-5 in a Likert scale). Fuzzy set can handle these kinds of databases by defining membership through probabilistic assessment of the membership of elements in a set (membership of importance/unimportance in this study). Generally, the fuzzy set maps the universe of discourse into to a set of real numbers which denote the membership of the universe of discourse elements in the set (Kumar, 2005). The 5-scale data are transferred to continuous data according to the membership function (1).

Let A_i is the ith fuzzy sets defined in the universe x_e by membership function $_{Ai}$: x_e → [0,1] and B is the fuzzy set defined in the universe y_e by membership function $_B$: y_e → [0,1]. x_e and y_e is the eth instance multi-item of input and output respectively. The measure of support for fuzzy rule $_{Ai}$→ $_B$ is defined as

$$\mathrm{Supp}(\ _{Ai} \rightarrow\ _{B}) = \frac{1}{N}\sum_{e=1}^{N} T(\ _{A}(x^e),\ _{B}(x^e))$$

and the confidence for fuzzy rule is defined as

$$\mathrm{Conf}(\ _{Ai} \rightarrow\ _{B}) = \frac{\sum_{e=1}^{N} T(\ _{A}(x^e),\ _{B}(x^e))}{\sum_{e=1}^{N} T(\ _{A}(x^e))},$$

where N is the size of database, $_{A}(x^e) = \min_i\ _{A_i}(x^e)$, *T* is the product T-norm.

2.4 Extracting Mining Rules by Genetic Algorithm

The Apriori algorithm is efficient in extract useful information from a large data transaction. However, several authors have pointed out some drawbacks of the support/confidence framework to assess association rules. Han et al. have pointed out that setting the mining support is subtle (Han et al., 2002). Also, the framework might lead to find misleading rules (Brin et al., 1997; Silverstein et al., 1998). To avoid the drawbacks and to ensure the discovered rules are interesting and accurate, a new approach was proposed (Silverstein et al., 1998; Delgado et al. 2003). The certainty factor is employed as the interestingness measure as follows (Delgado et al. 2003 and Yan et al.):

$$\text{CF(A}\rightarrow\text{B)=}\begin{cases}\dfrac{\text{Conf}(A\rightarrow B)\text{-Supp}(B)}{1\text{-supp}(C)}, \text{ if Conf}(A\rightarrow B) > \text{Supp}(B)\\ \dfrac{\text{Conf}(A\rightarrow B)\text{-Supp}(B)}{\text{supp}(B)}, \text{ if Conf}(A\rightarrow B) \leq \text{Supp}(B)\end{cases}.$$

The value of certainty factor is between -1 and 1. It is positive when the dependence between A and B is positive, 0 when A and B are independence and negative when negative dependence. We only consider positive dependence in this research. In addition, the strength of the association rule in both direction (i.e. A→B and B→A) are considered in this study.

The genetic algorithm is an optimization procedure inspired by natural evolution was introduced by Davis (Davis, 1987). The algorithm is a robust searching algorithm in optimal problems. The genetic algorithm is adopted to find the optimized of the certainly factor efficiently. Instead of the traditional support/confidence framework, the algorithm does not required user-specified minimal support and minimal confidence. Only the most interesting rules are returned automatically according the interestingness measure defined by the fitness function.

3. Data Collection

Evaluating MEC is a sequential stage process. It contains three steps (1) eliciting the most relevant attributes, (2) using laddering process to reveal the links between attributes to consequences and values and (3) deriving the hierarchical value map to express results from the ladders (Reynolds and Whitlark, 1995). Laddering is a very elaborative process as it involves asking respondents initially the features of products that they see. The interviewer then leads respondents to abstraction by asking why that feature is important. A sequence of concepts can then be linked in a „ladder". Collecting data (qualitative) through laddering process requires well-trained interviewers and is very time-consuming when larger scale data are desirable, such as over 100 cases (Reynolds and Gutman, 1988). As an alternative, Ter Hofstede et al. (1998) suggested a quantitative approach of data collection and analysis by using association pattern technique (APT) with log-linear models when larger scale data is desirable for making inferences. As this study is intended in making generalization a combination of qualitative and quantitative data collection process is used as explained below.

As the first step of data collection in this study, the relevant attributes and consequences for using organic rice products were obtained from collating Web-based and focus groups comments made by those who have tried organic rice products. List of Values (LOVs) developed by Kahle et al. (1986) and value related to LOHAS and religion was adopted to be used in the questionnaire later. Comments were collected and collated in late 2010. All items in the five ladders (attributes, functional consequence, social consequence, instrument value and terminal value) of MEC were incorporated in a survey questionnaire for respondents to rate on a 5-point Likert scale, ranging from very unimportant to very important. Potential respondents were intercepted at organic specialty shops. The strength of linkages between attributes and consequences, and between consequences and values were evaluated by using a two-stage process of fuzzy association mining rules, i.e., using fuzzy membership function to transform 5-scale Likert data into continuous data (i.e., probability of 0, 0.125, 0.500 ,0.875 and 1.0000 of implying importance levels which is discussed further) as this is a more detailed account of importance/unimportance.

4. Results

There were a total of 400 responses and 250 of them were usable. Of the 250 valid cases, about 74.7% of the respondents were female and 25.3% male. About 77.4% of the respondents were married. More than 95% of them had tertiary education or above (which is not surprising in Taiwan). Close to 70% of the respondents were above the age of 40. Close to 80% of the respondents was Buddhist.

Sixteen attributes, seven functional consequence, eleven social consequence, seven instrument value and fourteen terminal value were identified in the literature and deemed useful for this study. All of the elements in the five levels of hierarchy were included for evaluation of their strength of linkage.

As the break condition for evolution of genetic algorithm, our approach uses a simple rule: either when the prediction error is zero (optimal solution) or when a fixed number of generations have been evaluated (default values is 100 individuals per generation). Selection function of genes is set as normalized geometric distribution. The crossover and mutation operators are two basic operator of the genetic algorithm. In this study, the crossover is set as 0.8. And, the mutation rate is set as 1 over population size. The population size of each iteration step is 20. Figure 2 indicate the highest certainty factor of product attributes, functional consequences, and values.

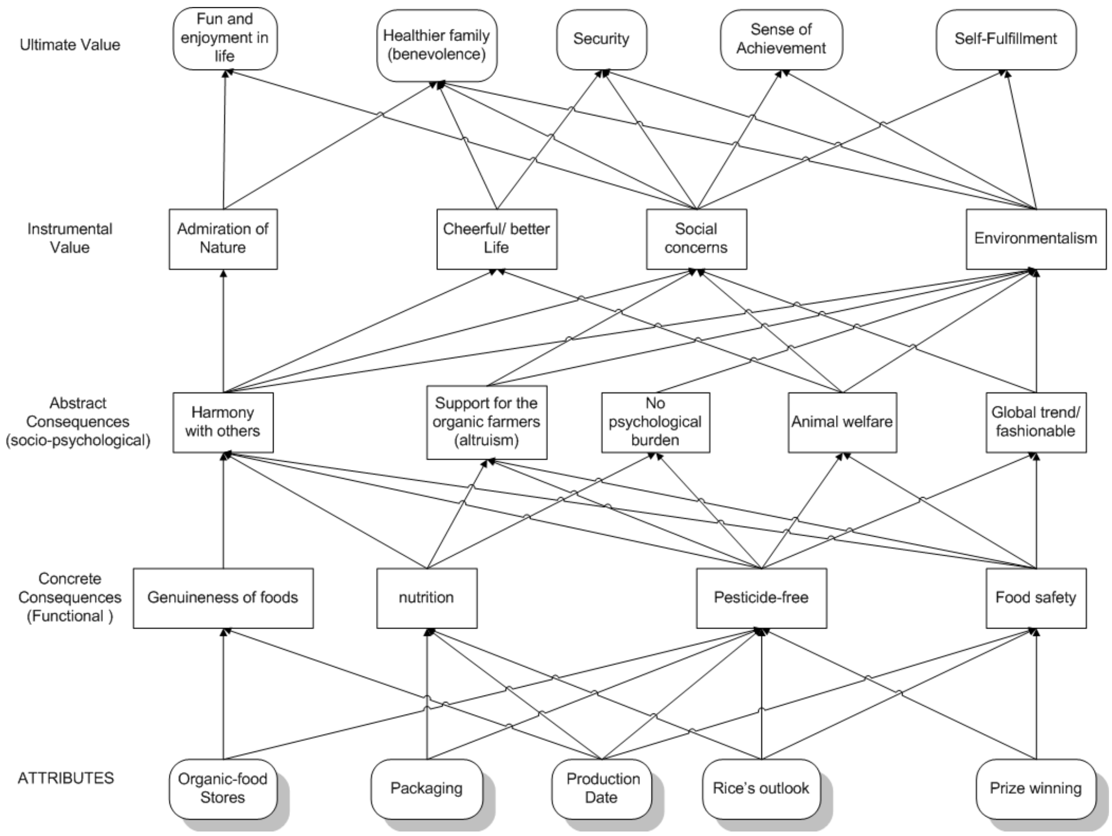

Figure 2. Means-End Chains of Organic Rice with Fuzzy Association Mining Method

5. Conclusions and discussions

This study applied means-end-chain framework to build a model for understanding buying motivations for organic rice consumers and potential consumers. Result shows that Taiwanese consumers concern about nutrition, genuineness of food taste, pesticide-free, and food safety, by paying more attentions to production date, packaging, prize winning and the channel stores than other attributes. Beyond these functional consequences and self interests, the green consumers also emphasize harmony with others, the animal welfare, giving supports to organic farmers, along with the global trend which green consumption is fashionable.

References

Agrawal, R. and Srikant, R. (1994), " Fast algorithm for mining association rules", Proceedings of the 20th International Conferences on Vary Large Data Bases, Santiago, Chile, pp. 487-499.

Beatty SE, Homer P, Kahle L.R. (1991), "Personal values and gift-giving behaviors: a study across cultures", Journal of Business Research, Vol.22, pp. 149-157.

Belk, R. (1988), "Possession and the Extended Self", Journal of Consumer Research, Vol.15, pp. 139-168.

Brin, S. Motwani, R. Ullman, J. D. and Tsur, S., (1997), "Dynamic itemset counting and implication rules for market basket data", SIGMOD Record, Vol. 26, No. 2, pp. 255- 264.

De Cock, M, Cornelis, C. and Kerre, E. E, (2005), "Elicitation of fuzzy association rules from positive and negative examples," Fuzzy Sets and Systems, Vol. 149, pp.73–85.

Davis, L., (1987), Genetic Algorithm and Simulation Annealing, Pitman publishing, London, 1987.

Delgado, M., Marĭn, N. Sănchez, D. and Vila, M. (2003), " Fuzzy association rules: General model and applications", IEEE Transactions on Fuzzy System, Vol. 11, No. 22, pp. 214-225.

Dubois, D. Prade, H. and Sudkamp, T. (2005), " On the representation, measurement, and discovery of fuzzy associations", IEEE Transaction on fuzzy systems, Vol. 13, No. 2, pp. 250-262.

Gutman, J. (1982), "A means-end chain model based on consumer categorization processes", Journal of Marketing, No.46, pp. 60-71.

Han, J., Wang, J., Lu, Y. and Tzvetkov, P., (2002), " Mining top-k frequent closed patterns without mining support", In Proceeding of the 2002 IEEE international conference on data mining (ICDM 2002), pp. 211-218.

Hong, T. P., Kuo, C. S. and Chi, S. C. ,(1999), " Mining association rules from quantitative data", Intelligent Data Analysis, Vol. 3, No. 5, pp. 363-340.

Kahle, L.R., Beatty, S.E., Homer, P. (1986), "Alternative measurement approaches to consumer values: the list of values (LOV) and values and life styles (VALS)", Journal of Consumer Research, Vol.13, pp.405-409.

Kotler, P. (2002), Marketing Management, 11th ed. NJ: Prentice-Hall, Inc.

Peter, J.P., Olson, J.C., Grunert, K. (1993), Consumer Behavior and Marketing Strategy, London: McGraw-Hill.

Pieters, R., Baumgartner, H., Allen, D. (1995), "A means-end chain approach to consumer goal structure", International Journal of Research Marketing, Vol.12, pp. 227-244.

Pitts, R.E, Wong, J.K., Whalen, D.J. (1991). "Consumers' evaluative structures in two ethical situations: a means-end approach", Journal of Business Research, Vol. 22, pp. 119-130.

Michell, A. (1983), The Nine American Lifestyles: Who We are and Where We"re Going. New York, MacMillan.

McColl-Kennedy, J.R., Kiel, G.C. (2000), Marketing: A Strategic Approach. South Melbourne, VIC: Nelson Thomson Learning.

Reynolds, T.J., Gutman, J. (1988), "Laddering theory, method, analysis, and interpretation", Journal of Advertising Research, Vol.1, pp.11-31.

Rokeach, M.J. (1968). Beliefs, Attitudes and Values. San Francisco: Jossey Bass.

Rokeach, M.J. (1973), The Nature of Human Values, New York: The Free Press.

Silverstein, C, Brin, S. and Motwani, R., (1998), "Beyond market basket: Generalizing association rules to dependence rules", Data Mining Knowledge Discovery, Vol. 2, pp. 39-68.

Soloman, Michael R. (2009), "The Self," in Consumer Behavior: Buying, Having and Being, 8th Ed.," Upper Saddle River, NJ: Pearson Prentice-Hall, pp.168-215.

Stanton, W., Etzel, J. and Michael, J. (1991). Fundamentals Marketing, 9th ed., London: McGraw-Hill.

Woodruff, R.B. (1997), "Customer value: The next source for competitive advantage", Academy of Marketing Science Journal, Vol.25, No.2, pp. 139-153.

Yan, X., Zhang, C. and Zhang, S., (2009), "Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support", Exper Systems with Applications, Vol. 36, pp. 3066-3076.

Yue, S., Tsang, E., Yeung, D. and Shi, D., (2000), "Mining fuzzy association rules with weighted items", The IEEE International Conference on Systems, Man and Cybernetics, 2000, pp. 1906–1911.

Zanoli R. and Naspetti, S. (2002), "Consumer motivations in the purchase of organic food", British Food Journal. Vol.104, pp.643-653.