Developing the Profiles of Super-Market Customers through Data Mining

Hokey Min

UPS Center for World Wide Supply Chain Management, University of Louisville, Louisville, KY 40292, USA
h0min001@gwise.louisville.edu

Abstract

To stay competitive, supermarkets need to develop a viable customer retention strategy. Since a key to the successful development of such a strategy rests with customer relationship management, supermarkets should identify the most profitable ways to build and maintain a loyal customer relationship. In an effort to help supermarkets understand their customers' shopping/patronage behavior and the ways to retain valued customers, we propose data mining techniques. Using the examples of franchised supermarkets in the Southeastern United States, this paper illustrates the usefulness of the proposed data mining techniques for examining customer grocery shopping behavior and developing the profiles of loyal patrons.

1. Introduction

Approximately half (51.1%) of U.S. households spent $75 or more for weekly grocery shopping in 1999 (David, 2000). However, supermarkets have historically operated on profit margins as low as one-half cent on every dollar of sales (Banning and Weber, 1994). With tight profit margins and increasing competition, a key to the supermarket's survival is its ability to cater supermarket services to the changing needs and life styles of ever-demanding grocery shoppers. Such needs include providing a greater selection of healthy foods, registered dietitians' free advice, fast-service checkout lines, better lighting, and lower shelves. Full understanding of such preferences, however, cannot be translated into a competitive advantage unless supermarket management develops detailed profiles of grocery shoppers. One of the most important purposes of customer profiling is to target "valued customers" for special treatment based on their anticipated future profitability to the supermarket. Customer profiling is the important basis for customer relationship management (CRM) and the subsequent development of customer retention strategy. In general, CRM is referred to as a business practice intended to improve service delivery, build social bonds with customers, and secure customer loyalty by nurturing long-term, mutually beneficial relationships with valued customers selected from a larger pool. As such, CRM often focuses on valued customers who repeatedly purchase a great amount of grocery items and remain committed to the particular supermarket over an extended period of time. In general, the longer customers remain with a particular supermarket, the more profitable they become to the supermarket (Reichheld and Sasser, 1990; Lovelock and Wright, 2002). Since these valued customers are more profitable to serve than others, the supermarket's customer retention efforts should be geared toward them. This paper proposes data mining techniques to answer the above questions. Data mining techniques are suitable for the profiling of grocery shoppers due to their proven ability to recognize and track patterns within a set of data

2. Research Methodology

The participants in this study were 301 grocery shoppers who have shopped at ten different supermarkets (Kroger, Meijer, Winn Dixie, A & P, Food Max, Food World, Pic-Pac, Aldi, Save-A-Lot, Piggly Wiggly) in the Southeastern United States during the period of 2001 through 2003. These supermarkets have been chosen for the study due to their similar characteristics in terms of franchising, location, and service amenities (availability of special departments and fast checkouts). Through a questionnaire survey, the participants provided us with data related to their demographic profile, how frequently they visit the supermarket, their shopping patterns, the relative importance of service attributes to overall supermarket service quality, and the level of customer satisfaction based on their service experiences. Some of the non-demographic questions were selected from service attributes critical to service quality (Lewis, 1987; Cadotte and Turgeon, 1988; Chung and Hahn, 1995; Min and Min, 1996; Vazquez et al., 2000; Min et al., 2002; Lovelock and Wright, 2002). Notice that customer satisfaction measures the perception of what actually happened in a service encounter compared with what the customer thought would happen in the situation (Zifko-Baliga, 1999).

All of the participants reported having visited at least one of the ten supermarkets for this study. In fact, a majority (72.7%) of the participants said that they had

visited one of these supermarkets at least once a week. A vast majority (95.7%) of them reported shopping at one of these supermarkets more than once a month. About three fourths (74.3 %) of them bought an average of more than ten items per visit. Most of the participants are frequent shoppers of the supermarkets and thus are knowledgeable about supermarket service quality.

In particular, these participants were carefully selected to maximize responses to the survey. Rather than distributing the questionnaire to the potential participants through the mail, seven surveyors (the author plus six hired graduate assistants) approached a group of grocery shoppers who had just checked out of the supermarket or handed over the questionnaire to the grocery shoppers through local church organizations and apartment complexes where the surveyors resided. Occasionally, the reluctant survey participants were offered a small financial incentive (a grocery coupon of no more than $5). In addition, a souvenir item with a nominal value was offered to the participants to encourage them to fill out the questionnaire.

Traditionally, data collected from questionnaires is analyzed using statistical techniques that aim to verify certain premises through a series of hypotheses testing. The examples of such traditional methods investigating service quality include Lewis (1987), Cadotte and Turgeon (1988), Chung and Hahn (1995), Stacey (1995), Min and Min (1996), Vazquez et al. (2000), and Min et al. (2002). Though still useful for certain situations, hypothesis testing could be inconsistent unless the level of statistical significance was decreased as the sample size increased (Glaymour et al., 1996). Hypothesis testing is also limited to either substantiating or disproving a pre-conceived notion. In other words, although hypothesis testing may allow us to validate some intuitive premises, causal inferences made by hypothesis testing may not be sufficient enough for us to accurately predict behavioral patterns of grocery shoppers in different supermarket settings. To overcome such a potential shortcoming, we employed data mining techniques.

3. Customer Profiling using Knowledge Discovery Rules

"IF-THEN" rules can give supermarket management clear customer profiles on which to target marketing or promotional actions. With this in mind, we developed more than fifty different rules (including sub-rules) using the *Clementine 8.0* Software (2003). Notice that the "IF-THEN" rules produced by a decision tree extract only useful information for formulating a customer retention strategy with a predictive accuracy surpassing 80%. Herein, the predictive accuracy represents a measure of a particular rule that reflects the proportionate number of times (percentage) that the rule is correct when applied to data. In other words, the predictive accuracy refers to a probability or confidence that the rule consequent ("THEN" part) holds true given that the rule antecedent ("IF" part) is present (see, e.g., Berry and Linoff, 2000; Han and Kamber, 2001). To produce a small, manageable set of rules, we initially extracted rules from training data sets (i.e., data objects whose class label is known) by setting a minimum threshold accuracy of 50% and then selected the rules whose predictive accuracy exceeds 80%. Considering the exponential complexity associated with the increased number of branches in the tree, we decided to list the selected sets of "IF-THEN" rules rather than constructing an entire decision tree (Groth, 2000).

"IF-THEN" rules summarized the grocery customer's shopping pattern with respect to the number of items that the customer purchased per visit, the frequency of visits to a supermarket, and the patronage of a particular supermarket. The volume of grocery purchases at the particular supermarket is influenced by the frequency of grocery shopping and the customer's loyalty (patronage) to that supermarket. To elaborate, the more frequent shopper tends to purchase more items per visit than does the less frequent shopper. Also, one who patronizes a particular supermarket tends to purchase more items per visit than one who does not. The marital status and gender of the customer influence his/her volume of grocery purchases. For example, either the married customer or female customer tends to purchase more items per visit than their single or male counterparts. Similarly, the age of a customer is related to his/her volume of grocery purchases In particular, customers in their twenties tend to purchase more items than other age groups.

In general, most unmarried customers tend to shop alone, whereas married but infrequent shoppers in the age group of 20-49 years old tend to shop with someone (e.g., their spouses). This implies that married shoppers tend to visit supermarkets together, when both couples are available for shopping (e.g., after working hours or weekends). Also, customers in their thirties who do not patronize a particular supermarket turned out be the most frequent shoppers. However, these customers seem to be more volatile and consequently more likely to defect to the competition than other customers of different age groups. Also, senior citizens who are older than sixty tend to shop more frequently than do other age

groups. On the other hand, customers in their fifties are the least frequent shoppers.

4. Major Findings and Implications

So far we have carried out a data mining study using the customer databases of ten supermarkets located in the Southeastern United States. One of the major objectives of this study was to classify the existing customer databases into certain types of segmentations and then predict a certain kind of customer behavior in selecting a particular supermarket, shopping frequency, purchase volume, and so forth. In addition, this study examined the influence of the customer's demographic profile (age, gender, marital status), frequency of supermarket visits, grocery purchase volume and degree of importance of certain service attributes, on the supermarket selection or patronage behavior. From a practical standpoint, several findings are noteworthy.

First, the customer's loyalty (patronage) to a particular supermarket and shopping frequency seem to influence his/her volume of grocery purchases. Certain customers' age and solitary/group shopping behavior also have some bearing on the volume of their grocery purchases. For instance, the patrons of a certain supermarket tend to buy more items per visit than do those who shopped at a number of different supermarkets. Therefore, supermarket management should enhance customer loyalty to increase potential revenue. Also, as expected, the more frequent shopper (e.g., those who shopped once a day) tends to buy a smaller number of items per visit than does the less frequent shopper (e.g., those who shopped once a month). To retain the more frequent shopper as a loyal customer, supermarket management should cater their service amenities (e.g., a greater availability of express checkouts, self-service checkouts) to the shopper who purchases a smaller number of items and may need faster checkout services.

Second, we found a systematic pattern between the marital status of the customer and his/her solitary or group shopping behavior. Congruent with common sense, a single (unmarried) customer tends to shop alone. This finding implies that a single customer is likely to make a supermarket patronage decision by himself/herself. Since some of the customers who shop alone tend to prefer fast checkout and easy payment, a supermarket that is located close to a neighborhood with a large percentage of unmarried customers (e.g., college campus area) may cater its service amenities to such customers by increasing the number of express checkout or self-service checkout counters and accepting various payment methods including credit cards, debit cards, and gift certificates. On the other hand, a supermarket that targets married individuals (especially in their twenties, thirties and forties) as their primary customer bases may offer a wide assortment of food and drinks that are catered to different needs of married couples. Also, it can extend store operating hours (e.g., weekends or late evening hours) during which both husbands and wives are available for shopping.

Third, we found a systematic pattern between the age of customers and their appreciation of employee courtesy. To elaborate, either very young (under 20) or older (over 50) customers tend to appreciate the courteous attitude of supermarket employees more than other age groups. The possible explanation for this pattern is that customers under 20 are relatively inexperienced shoppers and therefore may need greater assistance from supermarket employees in finding items that they want and validating their identifications for non-cash payment options. Similarly, those customers who are older than 50 are more likely to suffer from some physical impairment and subsequently need greater assistance from supermarket employees in reaching items on the taller shelves and picking up heavy items. Based on this observation, a supermarket that is in the proximity of retirement communities should be aware of the importance of employee courtesy to customer retention and consequently should enhance its employee training programs.

Fourth, we discovered that senior customers, who are older than 50, are very sensitive to the price of grocery items. This finding makes sense since many senior citizens (especially retirees) are more likely to rely on their limited retirement savings and government aids and therefore tend to be more price conscious than customers in their forties. Thus, a supermarket that is primarily targeting senior customers should consider offering greater money-saving opportunities (e.g., coupons, discounts for senior citizens, incremental discounts off the total purchase, customer shopping cards, general brand products) for them.

As illustrated above, a certain demographic has a distinctive pattern for patronizing a supermarket. However, our data mining results often indicate that a combination of multiple customer characteristics has greater influence on the customer's supermarket selection and patronage behavior than a single attribute. Finally, we found that different supermarkets tended to have different profiles of customer bases. In other words, some supermarkets have geared their services toward certain age groups. For example, relatively unknown, regional franchise supermarkets such as Pic Pac and

Aldi seem to attract younger patrons (e.g., in their 20s) more frequently than other age groups. Given their limited exposure and resources, their niche marketing strategy targeting younger customers makes sense. Therefore, their future customer retention strategy should focus on the enhancement of service amenities tailored toward younger customers.

Full tables, exhibits, and references will furnished from the author upon request.