Marketing Information System in Fitness Clubs
- Data Mining Approach -

Chen-Yueh Chen, Yi-Hsiu Lin, & David K. Stotlar
(University of Northern Colorado,USA)

Abstract

The fitness club industry has encountered a significant problem in high attrition rates. This paper attempted to utilize a data mining approach to identify the demographic characteristics of members who are likely to leave. The logistics regression was employed as a methodology. The input in this study was the information extracted from the membership database of a moderately sized multi-service fitness center in Colorado, United States of America. The hit rate for the overall model was 82%. The demographic characteristics of members who are more likely to leave are as follows: not-delinquent individuals; those members who have lower frequencies of coming to the fitness center; temporary memberships; individuals with a service change; members who use a monthly payment method. In addition, the marketing implications linked to the results were also presented.

Introduction

The demand in fitness clubs continues to grow across the U.S. The number of people joining fitness centers increased from 17.4 million in 1987 to 33.8 million in 2001 (American Sports Data, 2002). Over the past decade the prospects of this industry looked optimistic. However, the attrition rate in the fitness industry is also very high. According to the International Health, Racquet and Sport sclub Association (IHRSA, 2004), the membership attrition rate in this industry from 2001 to 2002 was approximately 40%. Yet, "if fitness centers could lower the attrition rates by just one percent, they could save a substantial amount of money" (Proctor, 2002, p.16). In the current market, many fitness club managers focus on acquiring new memberships and neglect the importance of retaining their existing members. However, although "new member sales are important, retention efforts are imperative" (Scudder, 2003, p.26). Therefore, understanding how to retain members in fitness clubs has become a critical issue.

A number of articles have presented the ideas and administrative methods regarding member retention. They are as follows: maintaining the facility properly, only hiring staff who care about membership (Bentkowski, 2003); focusing on staff training, programs, and interaction with members (Kleinberg, 2003); "creating the right first impression of the fitness club" (Smith, 2004, p.30); giving new members personal recognition and professional promotion (Coffman, 2004); and creating member positive experience because "the quality of a member's experience is a key factor in retention" (Lang and Lundberg, 2003, p. 20). Although many of these methods have been exploited to maintain membership retention by some fitness clubs, the attrition rates in this industry were still high. Namely, there may be problems with implementing these tactics of retaining members, e.g., fitness centers fail to identify the demographics of the members who have a tendency to leave and fail to customize services to meet those needs.

If fitness centers can identify the characteristics of the members who are inclined to leave by utilizing a scientific method, they can better communicate with members and develop tailored tactics to meet their needs. As a consequence, this paper attempts to employ a quantitative framework to identify the demographic characteristics of members who are very likely to leave by analyzing a membership database of a fitness club in Colorado. No further discussions regarding why members leave will be made.

Quantifying the likelihood of members' leaving by analyzing members' actual behavior (stay or leave) along with demographic information in the membership database was the object of this paper. This paper is organized as follows. First, the brief literature review on applications of data mining is presented, followed by the methodology used in this study. The last two sections are a summary of the results and the conclusions.

Literature Review

Due to the advancement of Information Technology, building a customer database in order to improve the performance of business operations is feasible. Amazon online bookstore exemplifies a successful case using database marketing to provide its customers with customized services to meet its customers' needs. One salient example is that if you would like to buy or search some book on Amazon's website, its website provides you with a list of books that others who bought the book in question also purchased. This has been achieved by utilizing a customer database, i.e., by recording the purchase records of online customers, doing analyses with the data collected, and then giving customized recommendations to customers. Data mining has not only been adopted by many companies, but has also caught many researchers' attention. Even though "database marketing has become one of the most significant developments in the field of marketing in recent years, (Mitchell, 2003, p. 219)" no articles that used this approach were found in the major sports-management-related journals. Consequently, the data mining approach utilized in the field of sport warrants investigation.

The popularity and attention of data mining exists because companies that make use of it are more likely to accomplish the goals of cross selling and member retention (Lau, Chow, and Liu, 2004). In addition, data mining is a more efficient approach to targeting that requires fewer resources (Lau, Chow, and Liu, 2004). Corporations can employ the information extracted from customer database to "make the right offer to the right customer at the right time, thereby increasing sales and pleasing the customers" (Hughes, 2002, p. 102). The above statements define the reasons why data mining has obtained comprehensive application in many businesses. Unfortunately, this approach has not been commonly applied to the sport-related industry, especially to membership retention in fitness centers, even though most fitness centers or professional sport teams have established their own membership databases. It is beneficial and lucrative for fitness centers if they can extract relevant and profitable information from their membership databases. Hence, this paper attempts to scientifically identify the demographic characteristics of members who are most likely to leave by exploiting the data mining approach instead of using the mangers' or staff's by-chance judgments.

Methodology

Definitions of the Variables

This study focused on identifying the demographic characteristics of fitness club members who are likely to leave by calculating their probability of leaving. The dependent variable (Y) in this study, which was a categorical variable, was the current status of members in this fitness center, i.e., either active members or terminated members. With regard to data entry, active members were coded as 0, and terminated ones were coded as 1. The independent variables were as follows:

1. Whether members were delinquent individuals (X1): This showed how current members were with their payments on account. Due to the fact that this variable was a categorical variable, past dues members were coded as 1; not-past-due members were treated as 0.
2. Members' visit counts (X2): This represented the visits of each member per week.
3. Membership enrollment fees (X3): This indicated the amount of money members paid in the beginning of their enrollment.
4. Members' year-to-date sales amount (X4): This identified the cost of services purchased by members this year.
5. Membership type (X5, X6, X7, X8): Five membership types were included, such as regular membership, student membership, temporary membership, dependent and spouse membership, and other membership. Four dummy variables were employed to deal with this categorical variable. These five types of membership were coded as (1,0,0,0); (0,1,0,0); (0,0,1,0); (0,0,0,1); and (0,0,0,0), respectively.
6. Whether members were service change individuals (X9): This indicated whether members got charged a rebilling fee when account payment was past due. Those who got charged a rebilling fee when account was past due were coded as 1; those who did not were treated as 0.
7. Members' zip code (X10): This variable was included to distinguish members who lived in the site city from those who lived outside the site city. Members who lived in the site city were coded as 0; those who lived outside the site city were coded as 1.
8. Payment methods (X11): Two methods of payment for members, i.e., annual payment and monthly payment were employed. Annual payment was coded as 1; monthly payment was coded as 0.
9. Gender (X12): Male members were coded as 1; female members were coded as 0.

Data

The data set consisted of 5,544 members. After careful examination and deletion of incomplete data, only 4,304 members were retained for this study. Table 1 reports the descriptions of the data set. For examining the accuracy of this model, the whole data set was divided into 2 subgroups: an analysis sample and a validation sample (Malhotra, 2004). The analysis sample which was exploited to estimate the parameters in the model contained 4,089 subjects. On the other hand, the validation sample which was randomly selected from the whole data set to check the accuracy of the estimation obtained from the analysis sample included 215 subjects.

Table 1
Descriptions of the Data Variables

Independent Variables	Average	Number	Percent
Status of members			
Active		3,528	82
Terminated		776	18
Delinquent individuals or not			
Delinquent		2,842	66
Not delinquent		1,462	34
Membership type			
Regular		1,729	40.2
Student		181	4.2
Temporary		69	1.6
Dependent and spouse		1,670	38.8
Others		655	15.2
Service change or not			
Change		2,503	58.2
Not change		1,801	41.8
Members' zip code			
In site city		3,531	82
Outside of site city		773	18
Payment methods			
Monthly		4,183	97
Annual		121	3
Gender			
Male		2,304	53.5
Female		2,000	46.5
Members' visit counts	2		
Enrollment fees	39		
YTD sales amount	255		

Table 1 reveals the following characteristics of the data set being studied. Among all subjects, approximately 82% were active members; 18% were terminated members. Sixty six percent were delinquent individuals; 34% were not delinquent ones. The averages of the members' visit counts per week, enrollment fee, and YTD sales amount are two times per week, $39; and $255, respectively. Among five types of membership, regular and dependent and spouse membership were 80%. Fifty-eight percent of all the subjects were individuals with service change. In addition, most subjects lived in the site city (82%) and preferred monthly payments (97%). The proportion of male and female subjects was approximately equal.

Statistical Technique

Logistic regression is a regression model dealing with the problems which predict a discrete outcome, such as the group membership, by using independent variables that may be discrete or continuous (Tabachnick and Fidell, 2001, p. 517). That is, the response variable is a binary response variable, e.g., stay or do not stay in the fitness center. The emphases on the probability of a particular result as well as without assumptions regarding the distributions of the independent variables are the advantages of the logistic regression (Tabachnick and Fidell, 2001, p. 517). The logistic regression has been applied to a number of fields, such as medical and marketing research. Therefore, this study will use it to analyze data. All data processing was accomplished by utilizing SPSS 12.0.

Results

Results of Parameter Estimation

Table 2 reports the results of parameter estimation. Independent variables, including whether members were delinquent individuals, members' visit counts, membership type, whether members were service change individuals, and payment methods were statistically significant ($p< .05$). That means these statistically significant coefficients have a substantial impact on the probability of leaving for each member. In other words, members with different characteristics (independent variables) will have a different probability of leaving.

Table 2

Summary of Parameter Estimation (N = 4,089)

	B	SE B	p-value
Intercept	0.45	.13	.00*
X1	-1.31	.20	.00*
X2	-0.81	.08	.00*
X3	0.00	.00	.18
X4	0.00	.00	.33
X5	-2.53	.23	.00*
X6	0.43	.27	.10
X7	1.91	.37	.00*
X8	-1.83	.13	.00*
X9	-0.30	.12	.01*
X10	-0.19	.12	.13
X11	-1.60	.58	.00*
X12	-0.14	.10	.14

* $p< .05$.

The explanations of the change in the probability of leaving (shown in Table 3) are as follows based on Allison's (1999) calculations. On average, the probability of leaving was .19 lower if the member was a delinquent individual compared with a not-delinquent individual, .12 lower for a 1-unit increase in members' visit counts, .37 lower if the subject subscribed to regular membership rather than other membership. The probability of leaving was .28 higher if the subject belonged to the temporary membership category compared to the other membership category, .27 higher if the subject had dependent/spouse Membership rather than other membership. The probability of leaving was .05 higher if the member was a service change individual compared with a not-service-change individual and .24 higher if members used monthly payment compared with those employing annual payment.

Table 3

Change in the Probability of Leaving for Significant Model Coefficients

Significant independent variables	Change in the probability of leaving
Delinquent individuals or not	-.19
Members' visit counts	-.12
Regular membership	-.37
Temporary membership	.28
Dependent and spouse membership	.27
Service change individuals or not	.05
Payment methods	.24

Results of Prediction

This section focused on using the estimates from the analysis sample and the members' information from the validation sample to predict the accuracy of this model. In this study, the cutoff point for the probability of leaving was set .5. That is, the members with the predicted probability greater than .5 were categorized as members who were likely to leave. On the other hand, those with a predicted probability less than .5 were thought of as members who were likely to stay. The results of categorizing members as leaving members or staying members were compared with their actual behavior (leaving or staying). By doing so, the hit rate—the proportion of members who were correctly predicted to the overall subjects in the validation sample—could be computed to examine the

accuracy of this model.
Table 4 reveals the results of prediction. The overall percentage of correct prediction was 82.32%, which was the proportion of the sum of subjects who were terminated members and were predicted as terminated ones (25) and those who were active members and were predicted as active ones (152) to the total number in the validation sample (215). If the inspection focuses on each category, the hit rates for subjects who were terminated members and were predicted as terminated ones as well as those who were active members and were predicted as active ones are 62.5% and 86.86%, respectively.

Table 4
Results of Prediction

	Predicted behavior		
Observed status	Leave	Stay	Predict percentage
Leave	25	15	62.50
Stay	23	152	86.86
Overall percentage			82.32

Discussions and Conclusions

A stronger link between the results and marketing strategy should be established after obtaining the analytical outcomes (Tapp, 2002). The results in this paper show that the hit rate of overall model was 82.32%; the hit rates for terminated members and active ones were 62.5% and 86.86%, respectively, which is satisfactory. This satisfactory hit rate implies that this model could be used to assist fitness center managers to identify members who are likely to leave and its predictive result is better than by-chance guesses. In addition, fitness center managers can identify the demographic characteristics of members who are likely to leave: not-delinquent individuals are more likely to leave; members who have lower frequencies of coming to the fitness center are more likely to leave than those who come infrequently; regular memberships are less likely to leave, while temporary memberships are more inclined to leave; a service change individual is more likely to leave than a not-service-change individual; members who use monthly payment method are more likely to leave than those who use an annual payment method. After identifying the characteristics of members who are more likely to leave, fitness center managers can identify who are likely to leave based on such information and tailor tactics to meet their needs and retain them. In other words, once fitness club managers understand the demographics of members who are likely to leave, it is beneficial to develop programs that engage in membership retention (Talmage, 2004). In this study, the characteristics of members who were likely to leave were presented by utilizing the database approach as well as the logistic regression. Customized marketing efforts should be designed to engage in membership retention. One thing that should be kept in mind by fitness club managers is that updating the information for each member on a regular basis is critical. Additionally, exploiting the information of the database can assist managers in providing members with the most relevant services as a more effective and efficient way to raise membership retention rates (Lau, Wong, Ma and Liu, 2003).
In conclusion, the utilization of the logistic regression with an extensive customer database was found to be a viable tool in identifying the demographic characteristics of members who are likely to leave. Then, employing the information extracted from the analyses to design member-retention tactics reduces member attrition in the health and fitness industry. While this was an introductory test of this method, the multiplications for use in other sports settings are clear. Hopefully, the results of this study can be applied to a variety of other marketing and sales situations across the sport and leisure industry.

References

Allison, P. D. (1999). Logistic regression using the SAS system: Theory and application. Cary, NC: SAS Institute Inc.
American Sports Data, Inc., (2002). The Superstudy of Sports Participation.
Bentkowski, F. (2003). Service for retention. Club Industry, 19(2), 22.
Coffman, S. (2004). Retention has never been more important than now. Club Industry, 20(2), 17.
Hughes, A. M. (2002). Editorial: The mirage of CRM. Journal of Database Marketing, 9(2), 102-104.
Membership growth and attrition. (2004). International Health, Racquet and Sportsclub Association. Retrieved from September 20, 2004, from

http://cms.ihrsa.org/IHRSA/ViewPage.cfm?pageid=809
Kleinberg, L. (2003). The retention plan. Club Industry, 19(4), 26.
Lang, A. and Lundberg, S. (2003). Personal experience. Club Industry, 19(2), 20.
Lau, K., Wong, S., Ma, M., and Liu, C. (2003). 'Next product to offer' for bank marketers. Journal of Database Marketing, 10(4), 353-368.
Lau, K., Chow, H., and Liu, C. (2004). A database approach to cross selling in the banking industry: Practice, strategies and challenges. Database Marketing & Customer Strategy Mnagement. 11(3), 216-234.
Malhotra, N. K. (2004). Marketing research: An applied orientation 4th ed. NJ: Pearson Education, Inc.
Mitchell, S. (2003). The new age of direct marketing. Journal of Database Marketing, 10(3), 219-229.
Proctor, L.D. (2002). Fostering member success. Club Industry, 18(10), 16.
Scudder, M. S. (2003). Retention by letters. Club Industry, 19(1), 26.
Smith, B. (2004). No second chances at first impressions. Club Industry, 20(1), 30.
Tabachnick, B. G., and Fidell, L. S. (2001). Using multivariate statistics 4th ed. MA: Allyn & Bacon Inc.
Talmage, D. (2004). Go beyond stage one when it comes to retention. Club Industry, 20(7), 29.
Tapp, A. (2002). The strategic value of direct marketing: Expanding its role within the company, paper 2. Journal of Database Marketing, 9(2), 105-112.