Inf Syst E-Bus Manage (2015) 13:475–494
DOI 10.1007/s10257-014-0264-1

ORIGINAL ARTICLE

Predicting customer churn from valuable B2B customers in the logistics industry: a case study

Kuanchin Chen · Ya-Han Hu · Yi-Cheng Hsieh

Received: 15 October 2013 / Revised: 28 June 2014 / Accepted: 22 September 2014 /
Published online: 2 October 2014

Abstract This study uncovers the effect of the length, recency, frequency, monetary, and profit (LRFMP) customer value model in a logistics company to predict customer churn. This unique context has useful business implications compared to the main stream customer churn studies where individual customers (rather than business customers) are the main focus. Our results show the five LRFMP variables had a varying effect on customer churn. Specifically length, recency and monetary variables had a significant effect on churn, while the frequency variable only became a top predictor when the variability of the first three variables was limited. The profit variable had never become a significant predictor. Certain other behavioral variables (such as time between transactions) also had an effect on churn. The resulting set of predictors of churn expands the original LRFMP and RFM models with additional insights. Managerial implications were provided.

Keywords Customer churn · Logistics industry · Customer value analysis · Prediction model

1 Introduction

Logistics is the flow of raw materials or other goods to end customers (Waters 2003; Yildiz et al. 2010). Delivering items to the correct place at an appropriate time at a reasonable cost is an essential success factor in the logistics industry (Liu and Lyons

K. Chen
Department of Business Information Systems, Western Michigan University,
3344 Schneider Hall, Kalamazoo, MI 49008-5412, USA

Y.-H. Hu (✉) · Y.-C. Hsieh
Department of Information Management, National Chung Cheng University,
Chiayi 62102, Taiwan, ROC
e-mail: yahan.hu@mis.ccu.edu.tw

2011). Concern arises when any stage of this process causes customer dissatisfaction, which potentially leads to loss of business. Logistics companies currently engage in activities that transcend the traditional goods delivery model, such as service delivery, rendering, and integration (Esper et al. 2003; Renko and Ficko 2010). For example, UPS has expanded its business, serving as a solution provider that offers service integration in e-commerce, accounts receivable, insurance, and financing. These activities relate to UPS's core business, and such expansion into other areas develops customer dependency and loyalty, which eventually translate into sustainable business.

A long-term relationship with customers is crucial in the logistics industry because numerous aspects of service encounters (such as cost, speed of delivery, and courtesy) can easily be imitated by competitors. Customer churn (or customer attrition) is key for gauging success in the logistics industry. Van den Poel and Lariviere (2004) showed that the cost of a high churn rate (at 25 % compared with 7 %) for a UK retail bank was nearly 220,000 euros after 25 years. Although this figure varies among industries, a highly competitive market, such as the logistics industry, is unlikely to have a low customer churn rate. Therefore, researchers recommend a defensive marketing strategy (Fornell and Wernerfelt 1987; Ahn et al. 2006) that prevents customers from switching service providers. Customer churn prediction is a popular element for retention and loyalty reasons (De Bock and Van den Poel 2012; Glady et al. 2009; Huang et al. 2012; Hung et al. 2006; Kisioglu and Topcu 2011; Li et al. 2011b; Neslin et al. 2006).

The purpose of customer churn management is to minimize losses caused by customer churn and to retain high-value customers, thereby maximizing profit. The 80/20 rule suggests that 80 % of revenue is provided by 20 % of customers (Xu and Walton 2005). Focusing on retaining high-value customers is a reasonable strategy. Consequently, studying the profile of lost customers and predicting customer churn is crucial for survival in highly competitive industries. Customer churn prediction has been applied in numerous fields, particularly in the telecommunication and financial industries, in which target audiences are primarily individuals (Huang et al. 2012; Verbeke et al. 2012; Kisioglu and Topcu 2011; Nie et al. 2011; Tsai and Lu 2009). Most high-value customers in the logistics industry are businesses that are bound by regulations, corporate policy, and accumulated business practice. Because of this difference, switching cost is considerably lower for individual customers than for business customers.

The contribution of this study is twofold. First, we expanded the body of knowledge in customer value analysis, specifically, the length, recency, frequency, monetary, and profit (LRFMP) model, to study the effect of such a model on business customers in the logistics industry (Li et al. 2011a; Verbeke et al. 2012). The results indicated that this model is generalizable in the logistics industry business context. Second, we investigated treating the effect of all five LRFMP variables concurrently. If some LRFMP variables overshadowed other variables, then further insight may be attained regarding how variables become predictors when variability of primary variables is limited or controlled.

2 Related work

2.1 Customer value analysis

Customer value analysis involves identifying patterns and group associations by using a high number of customer data. By conducting customer value analysis, businesses can identify valued customers who have contributed to business revenue the most for a considerable period of time (Cheng and Chen 2009). Among techniques for assessing customer value, the recency, frequency, and monetary (RFM) model has received considerable attention in recent literature (Chang et al. 2010; Chen et al. 2005; Cheng and Chen 2009; Hosseini et al. 2010; Liang 2010; Liu and Shih 2005; Yeh et al. 2009). The RFM technique is used widely in customer behavior analysis to study customer value and market segmentation.

Hughes (1994, 2005) studied numerous historical transaction data and determined that the following three key behavioral indicators are relevant to customer value analysis.

1. Recency refers to the period of time between the previous purchase date and a set date (determined before analysis begins). The likelihood of repurchase is high when the duration between these two dates is short.
2. Frequency refers to the total number of purchases within a particular period of time. It is used to measure the interaction frequency between a customer and the business. A high level of interaction indicates customer loyalty.
3. Monetary refers to the total dollar amount of a customer's purchases within a particular period of time. It can be used to measure the contribution of a customer to revenue. The greater the amount spent on purchases is, the more the customer contributes to revenue.

The importance of these three indicators varies among industries. Identifying the relative weight of these indicators for the specific domain of business is useful. Studies have recently begun expanding the RFM model by including additional variables. For example, Wei et al. (2012) suggested that the longevity of a relationship with a customer affects customer loyalty. They suggested an LRFM model including the longevity of a relationship, a CRFM model in which the RFM model is applied to various product categories, and a CLVRFM model based on traditional RFM analysis.

Chang and Tsai (2011) added product category groups to develop the GRFM model. Yeh et al. (2009) developed a RFMTC model by including the first purchase time (*T*) and customer churn probability (*C*). These extensions to the traditional RFM model have generated mixed results.

2.2 Customer churn

To retain customers, companies engage in activities to satisfy customers and reduce customer defection. Customer retention is the ultimate goal of customer relationship management (Payne and Frow 2005; Reinartz et al. 2004). In increasingly competitive business environments, acquiring new customers is increasingly expensive and difficult. Retaining existing customers is a popular strategy that is comparatively less costly than attracting new customers (Reinartz and Kumar 2003). Several studies have shown that acquiring a new customer is usually five to six times more expensive than retaining a customer (Athanassopoulos 2000; Slater and Narver 2000).

The scenario in which customers cease transacting with a company is called customer churn or customer attrition (Neslin et al. 2006; Yu et al. 2011). Customers who end a relationship with a company and develop a new relationship with a different company are called "churners" (Kisioglu and Topcu 2011).

Customer churn causes revenue loss and other negative effects on corporate operations (Saradhi and Palshikar 2011). Therefore, establishing an accurate customer churn prediction model for identifying key factors that cause churn is crucial. Several key recent studies on customer churn are summarized in Table 1.

According to Table 1, most previous studies on customer churn prediction have focused on the banking, retail, and telecommunication industries. According to a review of relevant literature, no study has investigated customer churn prediction in the logistics industry. In addition, variables that affect churn vary substantially among industries. For example, average minutes of usage, age, and place of residence, which are used in the telecommunication industry, are not relevant to logistics (Kisioglu and Topcu 2011). Similarly, debt ratio, home ownership, and credit risk, which are applied in the banking industry, are not relevant to logistics (Burez and Van den Poel 2008). One method for gaining insight into customer churn in an industry on which little empirical guidance has been provided in the literature is to apply a common set of classification techniques derived from previous churn studies that used a set of variables relevant to the industry. The immediate benefit of using common techniques is that it enables researchers to provide new evidence regarding the generalizability of existing methods.

Table 1 shows a common set of techniques used in recent churn studies. This set of techniques includes logistic regression (LGR), decision tree analysis (C4.5), artificial neural network (multilayer perceptron, MLP), and support vector machine (SVM). These four techniques were applied in this study.

3 Methods

3.1 Data collection and preprocessing

The company investigated in this study was founded in 1938 and is one of the largest logistics companies in Taiwan. It has approximately 2,500 transportation vehicles and over 100,000 business customers. In 2010, its total revenue was over TWD 8 billion (roughly USD 271 million). The original dataset comprised data on 106,747 business customers who engaged in over 210 million transactions between March, 2010 and August, 2012.

Table 1 Recent studies on customer churn prediction

Author (year)	Data source	Variable	Technique
Huang et al. (2012)	A telecom company in Ireland	7 variables	Decision tree; logistic regression; naive bayes; linear classifier; artificial neural network; support vector machine
Verbeke et al. (2012)	A telecom company in Europe	19 variables	Decision tree; artificial neural network; support vector machine
Miguéis et al. (2012)	An European retail company	RFM variables	Sequence mining; logistic regression
Saradhi and Palshikar (2011)	One unit of a large corporation for 2 years	25 variables for each employee	Decision tree; naive bayes; support vector machine; random forests; logistic regression
Kisioglu and Topcu (2011)	A telecom company in Turkey	23 variables (9 variables after data pre-processing)	Bayesian belief network
Kumar and Ravi (2008)	A credit card dataset in the bank	N/A	Decision tree; artificial neural network; support vector machine; random forest; logistic regression
Nie et al. (2011)	A commercial bank in China	135 variables	Decision tree; logistic regression
Huang et al. (2010)	A telecom company in Ireland	5 variable dimensions	Decision tree; artificial neural network; support vector machine; window techniques
Tsai and Chen (2010)	A wireless telecom company in Taiwan	22 variables	Decision tree; association rule; artificial neural network
Hung et al. (2006)	A wireless telecom company in Taiwan	10 variables	Decision tree; artificial neural network
Neslin et al. (2006)	A telecom company dataset	171 variables	Decision tree; logistic regression
Hwang et al. (2004)	A wireless telecom company in Korea	15 variables	Decision tree; logistic regression; artificial neural network
Wei and Chiu (2002)	A telecom company in Taiwan	9 variables	Decision tree; decision rule; artificial neural network

Before customer value analysis and customer churn prediction were conducted, a series of data preprocessing tasks, including merging customer, shipping, and delivery tables; removing records with missing values; deleting duplicate records; and aggregating records for each business customer, was performed. Moreover, recently acquired customers were excluded from the analysis because they had not been with the company long enough to be considered retained customers. These recent customers were defined as those for whom the transaction length (i.e., the number of days between the first and the final transaction) was shorter than 30 days. After recently acquired customers had been excluded, a total of 69,170 business customers remained in the final data set.

Before a customer churn prediction model was developed, the groups of active and lost customers were defined. In the case company, lost business customers were defined as those who engaged in no transactions in the past month. A customer service representative usually contacts lost customers to determine why they have not engaged in transactions. The case company classifies lost customers into four categories: those who changed location, those who went bankrupt, those who switched to a competitor, and those in debt. After consulting with experienced account managers, we considered only customers who switched to a competitor. The case company cannot address the attrition of customers in the other three categories. Among the 69,170 business customers, the numbers of active and lost customers were 67,849 and 1,321, respectively.

After consulting with the case company, we determined that a total of 18 variables (see Table 2) in three general categories were relevant to customer churn (customer profile, customer transaction behavior, and quality of delivery service). The descriptive statistics for both active and lost customers are shown in Table 3.

3.2 Experimental design

3.2.1 Customer value analysis

Figure 1 shows the research process, which can be divided into two main steps: customer value analysis and customer churn prediction. First, the differences in the scales of the LRFMP variables (see Table 2 for the scales) were standardized before further processing. The standardization procedure follows Hughes' (1994) equal depth approach (or called customer quintile method by Miglautsch 2000). Specifically, the customers were sorted in ascending order according to the variable CsnRcn (recency) and in descending order according to the other four variables. For each of the LRFMP variables, the customers were partitioned into equal quintiles. These quintiles were assigned numbers from 5 (highest customer value) to 1 (lowest customer value). This approach to standardizing scales has been the primary approach applied in other LRFMP studies (Cheng and Chen 2009; McCarty and Hastak 2007; Coussement et al. 2014).

To determine the weighting of each LRFMP variable, this study employed the analytic hierarchy process (AHP). Five senior sales managers at the case company were invited to evaluate the relative importance of the LRFMP variables. An example of the AHP pair-comparison matrix for the LRFMP model is shown in

Table 2 Definition of variables

Variable group	Variable	Type	Description
Customer profile	CsrRgn	Nominal	Customer region
	CsrNrRng	Numeric	Distance to the nearest branch (km)
Transactional behavior	CsnLng (length)	Numeric	Number of days between the first and the last transaction
	CsnRcn (recency)	Numeric	Number of days between the last transaction day and 2012-8-31
	CsnFrq (frequency)	Numeric	Average number of transactions per day
	CsnMnt (monetary)	Numeric	Average money spent per day
	CsnPft (profit)	Numeric	The ratio of high-delivery-fee transactions to total transactions
	CsnStpItvMin	Numeric	Minimum time interval between two adjacent transactions
	CsnStpItvMax	Numeric	Maximum time interval between two adjacent transactions
	CsnStpItvAvg	Numeric	Average time interval between two adjacent transactions
Quality of delivery	DlvErsNoth	Numeric	Number of failures of delivery/total transactions
	DlvErsApnt	Numeric	Number of redeliveries/total transactions
	DlvRchDay1	Numeric	Number of transactions delivered within 1 day/total transactions
	DlvRchDayOv3	Numeric	Number of transactions delivered over 3 days/total transactions
	DlvRchDayMin	Numeric	Minimum number of days for delivery
	DlvRchDayMax	Numeric	Maximum number of days for delivery
	DlvRchDayAvg	Numeric	Average number of days for delivery
Lost/active customer	CsrStRnEnt	Nominal	Lost or active customer

Table 3 Descriptive statistics of the complete customer data set

Variable	All customers (n = 69,170)	Lost customers (n = 1,321)	Active customers (n = 67,849)
CsrRgn [n (%)]			
North	29,594 (42.78 %)	691 (52.31 %)	28,903 (42.6 %)
Middle-north	9,352 (13.52 %)	117 (8.86 %)	9,235 (13.61 %)
Middle	14,862 (21.49 %)	347 (25.97 %)	14,515 (21.39 %)
Middle-south	7,564 (10.94 %)	103 (7.8 %)	7,461 (11 %)
South	7,220 (10.43 %)	60 (4.54 %)	7,160 (10.55 %)
East	578 (0.84 %)	3 (0.23 %)	575 (0.85 %)
CsrNrRng (mean ± SD)	3.189 ± 4.521	2.89 ± 5.104	3.195 ± 4.509
Length (mean ± SD)	679.0 ± 289.4	387.0 ± 242.5	684.7 ± 287.3
Recency (mean ± SD)	109.6 ± 200.5	420.5 ± 239.4	103.5 ± 194.8
Frequency (mean ± SD)	1.525 ± 34.485	1.213 ± 5.876	1.531 ± 34.81
Monetary (mean ± SD)	222.9 ± 2,403.3	155.5 ± 526.7	224.2 ± 2,425.5
Profit (mean ± SD)	0.393 ± 0.303	0.405 ± 0.316	0.393 ± 0.302
CsnStpItvMin (mean ± SD)	8.147 ± 46.90	9.444 ± 39.58	8.122 ± 47.03
CsnStpItvMax (mean ± SD)	92.51 ± 122.22	86.78 ± 101.43	92.62 ± 122.59
CsnStpItvAvg (mean ± SD)	24.17 ± 57.40	24.8 ± 46.61	24.16 ± 57.59
DlvErsNoth (mean ± SD)	0.028 ± 0.057	0.028 ± 0.063	0.028 ± 0.057
DlvErsApnt (mean ± SD)	0.013 ± 0.027	0.01 ± 0.026	0.013 ± 0.027
DlvRchDay1 (mean ± SD)	0.940 ± 0.071	0.938 ± 0.085	0.941 ± 0.071
DlvRchDayOv3 (mean ± SD)	0.009 ± 0.029	0.009 ± 0.038	0.009 ± 0.029
DlvRchDayMin (mean ± SD)	1.002 ± 0.171	1.002 ± 0.083	1.002 ± 0.172
DlvRchDayMax (mean ± SD)	4.722 ± 4.590	3.25 ± 3.36	4.751 ± 4.606
DlvRchDayAvg (mean ± SD)	1.068 ± 0.311	1.065 ± 0.319	1.068 ± 0.311

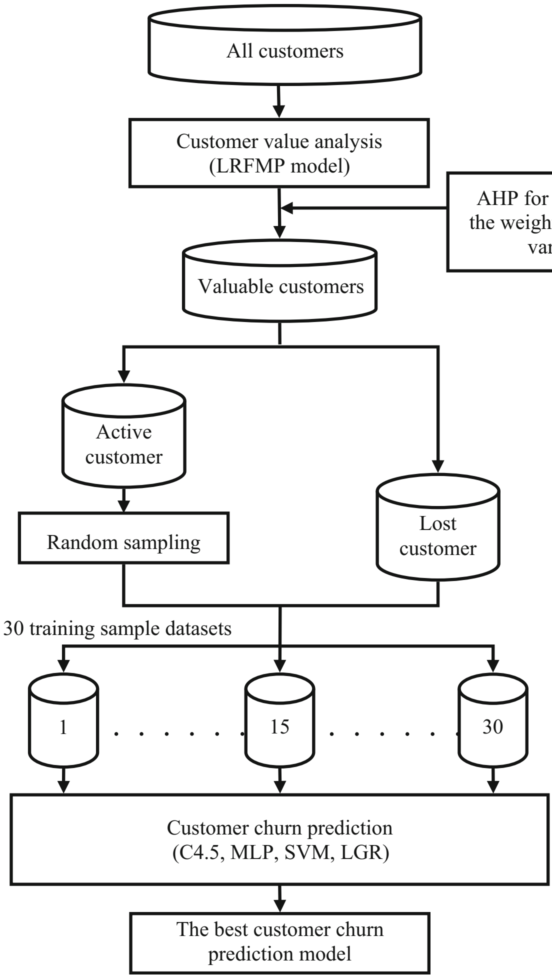

Fig. 1 Research process

Table 4. According to the AHP assessment, the relative LRFMP weights were 0.077, 0.225, 0.43, 0.241, and 0.026, indicating that frequency carries the most weight, followed by recency, monetary, length, and profit.

After the standardized LRFMP score for each customer was collected and the weight was determined for each of the LRFMP variables by using the AHP, a composite score for each customer was calculated as follows:

Assume that the standardized LRFMP scores of customer u_i ($u_i \in U$) are NL_i, NR_i, NF_i, NM_i, and NP_i, respectively. The LRFMP composite score of customer u_i, denoted as $Score_{ui}$, is defined in the following equation:

Table 4 An example of AHP pair-comparison matrix for the LRFMP model

Attribute	Comparative importance									Attribute
	9:1	7:1	5:1	3:1	1:1	1:3	1:5	1:7	1:9	
Length				✔						Recency
						✔				Frequency
								✔		Monetary
		✔								Profit
Recency							✔			Frequency
									✔	Monetary
			✔							Profit
Frequency						✔				Monetary
	✔									Profit
Monetary	✔									Profit

$$LRFMP(u_i) = 0.077 \times NL_i + 0.225 \times NR_i + 0.43 \times NF_i + 0.241 \times NM_i + 0.026 \times NP_i. \quad (1)$$

The coefficients of the variables in this equation were determined using the aforementioned AHP study. The composite scores were subjected to a median split to study patterns among the customers. The top 50 % of customers were labeled "valuable customers," whereas the remaining customers were labeled "less valuable customers." Median split is a common approach for studying patterns, behavioral difference, intention, and satisfaction in churn-related studies. Examples include word of mouth in customer lifetime value (Lee et al. 2006), and customer loyalty (Zhang et al. 2010).

3.2.2 Customer churn prediction

Because valuable customers contribute more to the profitability of the firm than do less valuable customers, we concentrated on these customers and their churn rate. We used Weka 3.7.7, a widely used open-source data mining software (www.cs.waikato.ac.nz/ml/weka), to study the performance of classification techniques, namely J48 (C4.5 in Weka), MultilayerPerceptron (MLP in Weka), SMO (SVM in Weka), and SimpleLogistic (LGR in Weka) (Linoff and Berry 2011; Tan et al. 2006).

The churn rate of valuable customers was approximately 1 %. Such a low customer churn rate can cause a class imbalance problem, wherein the majority class or group influences the prediction more than the minor class or group because of unequal representation. Previous studies have suggested that the sample size can be adjusted to improve the prediction performance of supervised learning (Tan et al. 2006). A resampling procedure was conducted to ensure that the lost/active ratio remained balanced. Specifically, we generated the dataset by undersampling the majority instances and retaining the complete set of minority instances so that the sample sizes of the two groups were approximately equal. In addition, useful instances in the majority class can be lost if the resampling procedure is applied only once. Therefore, a random resampling technique was applied 30 times to generate multiple datasets; for each generated dataset, tenfold cross-validation was applied to evaluate sample quality.

3.3 Parameter settings

According to the industry-wide data mining process model, the Cross Industry Standard Process for Data Mining, parameter calibration for the modeling phase of data mining involves testing the model by using a range of parameter values to optimize the model. Details on the parameter settings are shown in Table 5. Regarding C4.5, a decision tree stops growing if the number of instances in a node does not satisfy a user-specified threshold (i.e., 15, 20, or 25). Regarding MLP, we chose a single-hidden-layer network topology with a sigmoid function. Four other crucial parameters were adjusted in the experiments, including the number of nodes in the hidden layer, learning rate, momentum factor, and maximal number of epochs. In SVM, both the PolyKernel and RBFKernel were selected as the kernel function in all tests.

3.4 Performance measure

The performance of the classification models can be evaluated using several widely accepted indicators (accuracy, precision, recall, and F1) based on the confusion matrix. The confusion matrix in Table 6 can be used to calculate these four metrics as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN}. \tag{4}$$

$$F1 = \frac{2 \times Precision \cdot Recall}{Precision + Recall}. \tag{5}$$

4 Results and discussion

4.1 Results of customer value analysis

The purpose of customer value analysis is to identify valuable customers that potentially contribute to the profitability of the company. As previously discussed,

Table 5 Parameter tuning in different classification techniques

Method	Parameter	Range	Increment
C4.5	Minimum number of objects (O)	15–25	5
MLP	Number of hidden nodes (H)	10–13	3
	Learning rate (L)	0.1–0.3	0.2
	Momentum factor (M)	0.2–0.4	0.2
	Maximum number of epochs (N)	500–1,000	500
SVM	Kernel	PolyKernel	N/A
		RBFKernel	N/A
LGR	N/A	N/A	N/A

Table 6 Confusion matrix

Actual class	Predicted class	
	Lost customer	Active customer
Lost customer	*TP*	*FN*
Active customer	*FP*	*TN*

TP true positives, *TN* true negatives, *FN* false negatives, *FP* false positives

valued customers are the top 50 % of customers (a total of 34,675) based on the composite scores ($Score_{ui}$) calculated using the LRFMP variables. The churn rate for this group was approximately 1.1 %. The customers in this 1.1 % were lost customers, whereas the customers who remained with the company were active customers. Although a low churn rate in this group is a positive sign for a healthy future, it caused the problem of unequal representation of the two groups in data analysis. A resampling procedure was conducted to ensure that the sizes of the two groups were approximately equal. Table 7 shows descriptive statistics on LRFMP variables for the two customer groups. Table 8 shows how the two groups differed in other related variables.

4.2 Results of customer churn prediction

The data sets produced in the previous section were used to measure the performance of four classifiers, namely C4.5, MLP, SVM, and LGR. The results and accuracy measures are presented in Table 9. Accuracy, precision, recall, and the F1 measure were used to guide the direction of model tuning. To facilitate explanation, we report only the average values based on 30 generated datasets.

For each classification algorithm, we report only the optimal result (bold-faced numbers in Table 9) among various parameter settings. The highest accuracy rates for C4.5, MLP, SVM, and LGR were 93.1, 90.9, 88.4, and 87.6 %, respectively. The highest F1 measures of C4.5, MLP, SVM, and LGR were 93.3, 91.1, 87.2, and 86.8 %, respectively. A one-tailed paired *t* test was conducted to compare the

Table 7 Descriptive statistics of the complete customer data set in customer value analysis

Variable	All customers (n = 69,170)	Lost customers (n = 1,321)	Active customers (n = 67,849)
NL_i (mean ± SD)	3.301 ± 1.446	1.589 ± 0.712	3.059 ± 1.442
NR_i (mean ± SD)	2.744 ± 1.133	1.231 ± 0.487	2.774 ± 1.122
NF_i (mean ± SD)	3 ± 1.414	2.783 ± 1.402	3.004 ± 1.414
NM_i onetary (mean ± SD)	3 ± 1.414	2.763 ± 1.41	3.005 ± 1.414
NP_i (mean ± SD)	3.004 ± 1.413	3.041 ± 1.454	3.003 ± 1.412
$Score_{ui}$ (mean ± SD)	2.942 ± 1.157	2.341 ± 0.933	2.954 ± 1.158

prediction accuracy of the classifiers. The results indicated that C4.5 significantly outperformed the other three algorithms (C4.5 vs. MLP: $t = 22.4057$, $p < 0.001$; C4.5 vs. SVM: $t = 12.4619$, $p < 0.001$; C4.5 vs. LGR: $t = 34.2875$, $p < 0.001$). In addition, SVM and MLP significantly outperformed LGR (MLP vs. LGR: $t = 18.7484$, $p < 0.001$; SVM vs. LGR: $t = 2.144$, $p < 0.05$). Similar results were obtained regarding the F1 measure. C4.5 significantly outperformed the other three algorithms (C4.5 vs. MLP: $t = 18.7992$, $p < 0.001$; C4.5 vs. SVM: $t = 62.5325$, $p < 0.001$; C4.5 vs. LGR: $t = 36.2557$, $p < 0.001$). Furthermore, SVM and MLP significantly outperformed LGR (MLP vs. LGR: $t = 18.6723$, $p < 0.001$; SVM vs. LGR: $t = 3.063$, $p < 0.01$). These results indicated that C4.5 was the optimal algorithm according to all of the reported performance measures. LGR performed the least favorably. In addition, we observed that the sensitivity for parameter variation was low, indicating that all techniques used in the analysis are reliable classification models.

A customer churn prediction model can be used as an early warning tool for businesses, and extracting critical factors related to customer churn can provide additional useful knowledge that supports decision making. We conducted an additional experiment to identify the top three variables that exerted the most substantial effect on whether a customer is lost or active. The information gain ratio (or gain ratio) was used to determine the variables most relevant to the model. The results indicated that CsnRcn (days since the most recent transaction) was the most influential variable, followed by CsnLng (longevity of the relationship) and CsnMnt (average revenue gained from the customer). In other words, customers for whom the number of days since the previous transaction (CsnRcn) is higher, those who have maintained a shorter relationship with the company (CsnLng), and those who have provided lower average revenue to the case company (CsnMnt) have a higher probability of being a churner. This result was expected because most business customers have a constant need for the delivery of products or services. Unless a replacement delivery method is implemented (e.g., mail replaced by e-mail or electronic communications), the need for delivery does not fluctuate drastically in a short period of time. Therefore, a high value in either recency (CsnRcn) or longevity (CsnLng) is a possible indication of customer churn. This finding is consistent with those of previous studies, such as Chen et al. (2008) and Li et al. (2011a). The

Table 8 Descriptive statistics of the valuable customer data set

Variable	All customers (n = 34,675)	Lost customers (n = 409)	Active customers (n = 34,266)
CsrRgn [n (%)]			
North	15,166 (43.74 %)	213 (52.08 %)	14,953 (43.64 %)
Middle-north	4,409 (12.72 %)	42 (10.27 %)	4,367 (12.74 %)
Middle	7,673 (22.13 %)	89 (21.76 %)	7,584 (22.13 %)
Middle-south	3,789 (10.93 %)	38 (9.29 %)	3,751 (10.95 %)
South	3,456 (9.97 %)	25 (6.11 %)	3,431 (10.01 %)
East	182 (0.52 %)	2 (0.49 %)	180 (0.53 %)
CsrNrRng (mean ± SD)	3.162 ± 4.529	2.77 ± 4.57	3.167 ± 4.528
Length (mean ± SD)	778.38 ± 252.34	401.19 ± 254.61	782.89 ± 248.88
Recency (mean ± SD)	38.38 ± 131.3	421.89 ± 249.57	33.81 ± 122.18
Frequency (mean ± SD)	2.95 ± 48.67	3.65 ± 10.15	2.94 ± 48.94
Monetary (mean ± SD)	426.82 ± 3,382	452.43 ± 876.09	426.51 ± 3,400.79
Profit (mean ± SD)	0.371 ± 0.303	0.368 ± 0.321	0.371 ± 0.303
CsnStpItvMin (mean ± SD)	1.06 ± 2.56	1.049 ± 0.446	1.059 ± 2.573
CsnStpItvMax (mean ± SD)	35.19 ± 61.42	34.154 ± 59.385	35.203 ± 61.446
CsnStpItvAvg (mean ± SD)	3.62 ± 4.87	3.022 ± 3.305	3.63 ± 4.881
DlvErsNoth (mean ± SD)	0.026 ± 0.037	0.026 ± 0.026	0.026 ± 0.037
DlvErsApnt (mean ± SD)	0.013 ± 0.015	0.012 ± 0.015	0.013 ± 0.015
DlvRchDay1 (mean ± SD)	0.944 ± 0.04	0.944 ± 0.0.39	0.944 ± 0.04
DlvRchDayOv3 (mean ± SD)	0.008 ± 0.012	0.007 ± 0.011	0.008 ± 0.012
DlvRchDayMin (mean ± SD)	1 ± 0.054	1 ± 0	1 ± 0.054
DlvRchDayMax (mean ± SD)	6.893 ± 5.128	5.658 ± 4.228	6.908 ± 5.136
DlvRchDayAvg (mean ± SD)	1.058 ± 0.149	1.053 ± 0.07	1.058 ± 0.15

Table 9 Experimental results for all classifiers

Method	Parameter setting	Accuracy	Precision	Recall	F1
C4.5	O = 15	0.929	0.907	0.956	0.931
	O = 20	0.930	0.907	0.959	0.932
	O = 25	**0.931**	**0.907**	**0.961**	**0.933**
MLP	H = 10, L = 0.1, M = 0.2, N = 500	0.907	0.905	0.911	0.908
	H = 10, L = 0.1, M = 0.4, N = 500	0.904	0.900	0.910	0.905
	H = 10, L = 0.3, M = 0.2, N = 500	0.899	0.898	0.900	0.899
	H = 10, L = 0.3, M = 0.4, N = 500	0.896	0.897	0.894	0.896
	H = 13, L = 0.1, M = 0.2, N = 500	**0.909**	**0.904**	**0.914**	**0.911**
	H = 13, L = 0.1, M = 0.4, N = 500	0.906	0.902	0.912	0.907
	H = 13, L = 0.3, M = 0.2, N = 500	0.898	0.895	0.901	0.898
	H = 13, L = 0.3, M = 0.4, N = 500	0.896	0.895	0.898	0.896
	H = 10, L = 0.1, M = 0.2, N = 1,000	0.902	0.896	0.909	0.903
	H = 10, L = 0.1, M = 0.4, N = 1,000	0.900	0.892	0.910	0.901
	H = 10, L = 0.3, M = 0.2, N = 1,000	0.893	0.890	0.897	0.893
	H = 10, L = 0.3, M = 0.4, N = 1,000	0.892	0.890	0.894	0.892
	H = 13, L = 0.1, M = 0.2, N = 1,000	0.903	0.896	0.912	0.904
	H = 13, L = 0.1, M = 0.4, N = 1,000	0.900	0.891	0.912	0.901
	H = 13, L = 0.3, M = 0.2, N = 1,000	0.894	0.889	0.899	0.894
	H = 13, L = 0.3, M = 0.4, N = 1,000	0.891	0.888	0.907	0.9
SVM	PolyKernel	**0.884**	**0.933**	**0.819**	**0.872**
	RBFKernel	0.810	0.875	0.724	0.792
LGR	N/A	**0.876**	**0.923**	**0.820**	**0.868**

O, minimum number of objects; H, the number of hidden nodes; L, the learning rate; M, the momentum factor; N, the maximum number of epochs

Bolded numbers are the best results of the classification techniques

association of the longevity of the relationship (CsnLng) and average revenue provided to the case company (CsnMnt) with the target variables in the aforementioned experiments is consistent with the results reported by Buckinx and Van den Poel (2005). The consistency of these customer churn variables with those used in existing studies has numerous implications because the respondents were business users of delivery services. Such a combination of context and user profiles has not received sufficient attention in relevant literature.

4.3 Profile analysis

This section presents the results of profile analysis, of which the goal is to determine patterns between lost and active customers based on the similar magnitudes of the top three variables (CsnRcn, CsnLng, and CsnMnt) reported in the previous section. The data were first sorted according to CsnRcn (ascending order), CsnLng (descending order), and then CsnMnt (descending order). Because of this sorting

Table 10 Variable contribution to churn when variability of recency, length and CsnStpItvMax are limited

Rank	Gain ratio	Variable	Chi square	Variable
1	.03552	CsnRcn (recency)	714.0438	CsnRcn (recency)
2	.02659	CsnLng (length)	365.7026	CsnLng (length)
3	.02345	CsnStpItvMax	325.1557	CsnStpItvAvg
4	.01789	CsnFrq (frequency)	299.556	CsnFrq (frequency)
5	.01651	CsnStpItvAvg	238.6641	CsnStpItvMax

Table 11 The decision rules extracted by the C4.5

No.	Decision rule	
1	**IF** (CsnRcn > 148) **THEN** class = Churn	(359/24)
2	**IF** (17 < CsnRcn ≦ 148) & (CsnMnt > 52.3743) **THEN** class = Churn	(34.0/5.0)
3	**IF** (CsnRcn > 51) & (CsnStpItvMax > 63) **THEN** class = Churn	(49/10)
4	**IF** (CsnRcn > 20) & (CsnLng > 808) & (CsnStpItvMax ≦ 103) **THEN** class = Churn	(25/5)
5	**IF** (CsnRcn > 28) & (CsnLng > 793) & (DlvErsNoth > 0.0161) **THEN** class = Churn	(25/5)
6	**IF** (CsnRcn > 28) & (CsnLng ≦ 793) **THEN** class = Churn	(398/29)
7	**IF** (CsnRcn > 19) & (CsnStpItvMax ≦ 63) **THEN** class = Churn	(380/24)

procedure, the variability of records among these three variables in each quartile was limited. The top 50 % of this sorted data consisted of only 5 churners and 17,332 active customers, representing a churn rate of 0.03 % [5/(5 + 17,332)]; however, the bottom 50 % consisted of 404 churners and 16,934 active customers, representing a churn rate of 2.33 %. As these three variables were the top three predictors, churners were located mainly in the bottom portion of the sorted data set. Quartile 3 contained only 6 churners and 8,663 active customers. Therefore, most churners were located in Quartile 4, which contained 398 churners and 8,271 active customers, representing a churn rate of 4.59 %. This high churn rate is 153 times higher than that of the top 50 % of the sorted data set.

The next step involved explaining the top predictors of churn in the fourth quartile, wherein the variability of CsnRcn, CsnLng, and CsnMnt was limited because of the aforementioned sorting procedure. According to the gain ratio and Chi square statistic, the top predictors of churn in Quartile 4 were CsnRcn (recency), CsnLng (length), CsnStpItvMax, CsnFrq (frequency), and CsnStpItvAvg (as shown in Table 10). CsnMnt (monetary) did not qualify as a top predictor, whereas CsnFrq (frequency) did, indicating that the effect of frequency may have been shadowed until the variability of other variables was limited or controlled in the fourth quartile. In all analyses, profit was never influential. This insight improved our understanding of the relationship between the LRFMP variables and churn.

Regarding other profile variables, both customer region (CsrRgn) and distance to the nearest branch of the case company (CsrNrRng) exerted little effect on whether the customer was lost to competition. One possible reason is that the rural–urban disparity in Taiwan is low and the case company provides a free package pickup service for business customers. The statistics in Table 10 indicate that none of the quality variables were among the top predictors. Further analysis revealed that the case company provides high-quality delivery service according to several metrics. For example, the delivery failure rate was 0.026 %, the redelivery rate was 0.014 %, the rate of deliveries completed in 1 day was 94.1 %, and the average number of days for delivery was 1.06. These numbers indicate that the company provided excellent delivery service. In other words, delivery quality was not likely a key cause of customer churn.

Table 11 lists several crucial decision rules generated using C4.5. We report only some of the crucial rules for customer churn prediction because of space limitations. The first number at the end of each rule represents the number of customers satisfying the antecedent part of the rule, whereas the second number represents the number of customers incorrectly classified according to the rule. Decision tree induction can be used to characterize a group of churn customers, and the results can be applied in customer relationship management.

5 Conclusion

This study examined the variables that contribute to the customer churn of valuable customers at a logistics company. Valuable customers were identified based on their composite scores, which were calculated using the LRFMP variables. Because the overall ratio of lost customers to active customers was low, a resampling procedure was adopted to balance the representation of the two customer groups. Various prediction models were then used to compare the prediction performance. The effects of the LRFMP variables, as well as other profile variables, on customer churn were examined.

The decision tree model outperformed the other models (MLP, SVM, and LGR) in predicting customer churn. The top three most influential predictors for customer churn were recency (CsnRcn), length (CsnLng), and monetary (CsnMnt), whereas the other two LRFMP variables (i.e., frequency and profit) were not significant predictors; these results are not consistent with those of several other customer churn studies. This points to an issue that relates to whether the five LRFMP variables equally influence churn.

This study addressed how and when these other LRFMP variables become useful or influential. We limited the variability of the top three churn variables (i.e., recency, length, and monetary) by sorting the records based on the three variables and focusing on the quartile that contained the highest number of lost customers. The results indicated that recency (CsnRcn), length (CsnLng), the maximal time interval between two adjacent transactions (CsnStpItvMax), frequency (CsnFrq), and the average time interval between two adjacent transactions (CsnStpItvAvg) became predictors after the variability of the top three variables was limited. This result indicates that the effects of the five variables in the LRFMP model do not equally affect churn. Even in the quartile containing the customers that are most likely to churn (high recency, short length, and short duration between transactions), many customers still decided to remain with the company. Our identification of these “secondary” predictors, frequency, length, and time between adjacent transitions, provides insight into the intention to churn.

The primary contributions of this study are described as follows. First, our sample of business customers in the logistics industry represents a population that has not been previously explored. Many previous studies on churn have focused on individual users who usually have a lower switching cost than do business customers (Miguéis et al. 2012; Huang et al. 2012). Although insight into the logistics industry and churn of business customers is limited, studying the effect of LRFMP variables in this context provides evidence regarding the generalizability of the LRFMP and traditional RFM models. Second, variables related to churn vary greatly among industries. As indicated in Sect. 2, numerous variables used in other industries are either not relevant or not applicable (e.g., average minutes of usage, place of residence, debt ratio, home ownership, and credit risk). This study clarifies how churn can be predicted more accurately in the logistics industry.

Third, not all of the five LRFMP variables exerted an equal effect on churn. Previous studies (Wei et al. 2012) have laid the foundation for these five variables, whereas this study expanded on this foundation by showing that only recency, longevity, and monetary are the most influential predictors of the five variables. Frequency was not considered a crucial predictor until the variability of the aforementioned three variables was limited. These results revealed that frequency, monetary, and length are the variables that are the most related to churn for customers who are the most likely to leave the company. The profit variable of the LRFMP model never became a significant predictor in our analyses.

Several directions can be taken in the future. First, the data were provided by a single case company. Data from numerous companies can be collected and compared to enhance the generalizability of the LRFMP model further. Second, the LRFMP model is a popular customer value analysis model, but it is not the only model. As shown in the current study, not all five of the LRFMP variables exerted a notable influence on churn. Future studies can consider multiple customer value analysis models. Third, as mentioned in the Introduction, logistics companies have begun including other services, such as banking and accounts receivable, in their core business. Such an expansion increases customer “lock-in,” a common business practice used to retain customers. Future studies can report the effect of customer lock-in on customer churn.

Acknowledgments This research was supported by the National Science Council of the Republic of China under the Grant NSC 102-2410-H-194-104-MY2.

References

Ahn JH, Han SP, Lee YS (2006) Customer churn analysis: churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. Telecommun Policy 30:552–568

Athanassopoulos AD (2000) Customer satisfaction cues to support market segmentation and explain switching behavior. J Bus Res 47(3):191–207

Buckinx W, Van den Poel D (2005) Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. Eur J Oper Res 164(1):252–268

Burez J, Van den Poel D (2008) Separating financial from commercial customer churn: a modeling step towards resolving the conflict between the sales and credit department. Expert Syst Appl 35:497–514

Chang HC, Tsai HP (2011) Group RFM analysis as a novel framework to discover better customer consumption behavior. Expert Syst Appl 38(12):14499–14513

Chang EC, Huang SC, Wu HH (2010) Using K-means method and spectral clustering technique in an outfitter's value analysis. Qual Quant 44(4):807–815

Chen MC, Chiu AL, Chang HH (2005) Mining changes in customer behavior in retail marketing. Expert Syst Appl 28(4):773–781

Chen Y, Fu C, Zhu H (2008) A data mining approach to customer segment based on customer value. In: International conference on fuzzy systems and knowledge discovery, pp 513–517

Cheng CH, Chen YS (2009) Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst Appl 36(3, Part 1):4176–4184

Coussement K, Van den Bossche FAM, De Bock KW (2014) Data accuracy's impact on segmentation performance: benchmarking RFM analysis, logistic regression, and decision trees. J Bus Res 67(1):2751–2758

De Bock KW, Van den Poel D (2012) Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. Expert Syst Appl 39(8):6816–6826

Esper TL, Jensen TD, Turnipseed FL, Burton S (2003) The last mile: an examination of effects of online retail delivery strategies on consumers. J Bus Logist 24(2):177–203

Fornell C, Wernerfelt B (1987) Defensive marketing strategy by customer complaint management: a theoretical anlaysis. J Mark Res 24(4):337–346

Glady N, Baesens B, Croux C (2009) Modeling churn using customer lifetime value. Eur J Oper Res 197(1):402–411

Hosseini SMS, Maleki A, Gholamian MR (2010) Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Syst Appl 37(7):5259–5264

Huang BQ, Kechadi TM, Buckley B, Kiernan G, Keogh E, Rashid T (2010) A new feature set with new window techniques for customer churn prediction in land-line telecommunications. Expert Syst Appl 37(5):3657–3665

Huang B, Kechadi MT, Buckley B (2012) Customer churn prediction in telecommunications. Expert Syst Appl 39(1):1414–1425

Hughes AM (1994) Strategic database marketing. Probus Publishing Company, Chicago

Hughes AM (2005) Strategic database marketing. McGraw-Hill, New York City

Hung SY, Yen DC, Wang HY (2006) Applying data mining to telecom churn management. Expert Syst Appl 31(3):515–524

Hwang H, Jung T, Suh E (2004) An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. Expert Syst Appl 26(2):181–188

Kisioglu P, Topcu YI (2011) Applying Bayesian belief network approach to customer churn analysis: a case study on the telecom industry of Turkey. Expert Syst Appl 38(6):7151–7157

Kumar D, Ravi V (2008) Predicting credit card customer churn in banks using data mining. Int J Data Anal Tech Strateg 1(1):4–28

Lee J, Lee J, Feick L (2006) Incorporating word-of-mouth effects in estimating customer lifetime value. Database Mark Cust Strateg Manag 14(1):29–39

Li DC, Dai WL, Tseng WT (2011a) A two-stage clustering method to analyze customer characteristics to build discriminative customer management: a case of textile manufacturing business. Expert Syst Appl 38(6):7186–7191

Li Y, Deng Z, Qian Q, Xu R (2011b) Churn forecast based on two-step classification in security industry. Intell Inf Manag 3(4):160–165
Liang YH (2010) Integration of data mining technologies to analyze customer value for the automotive maintenance industry. Expert Syst Appl 37(12):7489–7496
Linoff GS, Berry MJ (2011) Data mining techniques: for marketing, sales, and customer relationship management. Wiley, Hoboken
Liu CL, Lyons AC (2011) An analysis of third-party logistics performance and service provision. Transp Res Part E Logist Transp Rev 47(4):547–570
Liu DR, Shih YY (2005) Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. J Syst Softw 77(2):181–191
McCarty JA, Hastak M (2007) Segmentation approaches in data-mining: a comparison of RFM, CHAID, and logistic regression. J Bus Res 60(6):656–662
Miglautsch JR (2000) Thoughts on RFM scoring. J Database Mark 8(1):67–72
Miguéis VL, Van den Poel D, Camanho AS, e Cunha JF (2012) Modeling partial customer churn: on the value of first product-category purchase sequences. Expert Syst Appl 39(12):11250–11256
Neslin S, Gupta S, Kamakura W, Lu J, Mason C (2006) Defection detection: improving predictive accuracy of customer churn models. J Mark Res 43(2):204–211
Nie G, Rowe W, Zhang L, Tian Y, Shi Y (2011) Credit card churn forecasting by logistic regression and decision tree. Expert Syst Appl 38(12):15273–15285
Payne A, Frow P (2005) A strategic framework for customer relationship management. J Mark 69(4):167–176
Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. J Mark 67(1):77–99
Reinartz W, Krafft M, Hoyer WD (2004) The customer relationship management process: its measurement and impact on performance. J Mark Res 41(3):293–305
Renko S, Ficko D (2010) New logistics technologies in improving customer value in retailing service. J Retail Consum Serv 17(3):216–223
Saradhi VV, Palshikar GK (2011) Employee churn prediction. Expert Syst Appl 38(3):1999–2006
Slater SF, Narver JC (2000) Intelligence generation and superior customer value. J Acad Mark Sci 28(1):120–127
Tan PN, Steinbach M, Kumar V (2006) Introduction to data mining. Addison Wesley, Boston
Tsai CF, Chen MY (2010) Variable selection by association rules for customer churn prediction of multimedia on demand. Expert Syst Appl 37(3):2006–2015
Tsai CF, Lu YH (2009) Customer churn prediction by hybrid neural networks. Expert Syst Appl 36(10):12547–12553
Van den Poel D, Larivière B (2004) Customer attrition analysis for financial services using proportional hazard models. Eur J Oper Res 157(1):196–217
Verbeke W, Dejaeger K, Martens D, Hur J, Baesens B (2012) New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. Eur J Oper Res 218(1):211–229
Waters D (2003) Logistics: an introduction to supply chain management. Palgrave Macmillan, New York City
Wei CP, Chiu IT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert Syst Appl 23(2):103–112
Wei JT, Lin SY, Weng CC, Wu HH (2012) A case study of applying LRFM model in market segmentation of a children's dental clinic. Expert Syst Appl 39(5):5529–5533
Xu M, Walton J (2005) Gaining customer knowledge through analytical CRM. Ind Manag Data Syst 105(7):955–971
Yeh IC, Yang KJ, Ting TM (2009) Knowledge discovery on RFM model using Bernoulli sequence. Expert Syst Appl 36(3, Part 1):5866–5871
Yildiz H, Ravi R, Fairey W (2010) Integrated optimization of customer and supplier logistics at Robert Bosch LLC. Eur J Oper Res 207(1):456–464
Yu X, Guo S, Guo J, Huang X (2011) An extended support vector machine forecasting framework for customer churn in e-commerce. Expert Syst Appl 38(3):1425–1430
Zhang JQ, Dixit AA, Friedmann RR (2010) Customer loyalty and lifetime value: an empirical investigation of consumer packaged goods. J Mark Theory Pract 18(2):127–140