Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2010 Proceedings

Americas Conference on Information Systems (AMCIS)

8-2010

Toward Optimal Churn Management: A Partial Least Square (PLS) Model

Hyunbo Cho
POSTECH, hcho@postech.ac.kr

Yeonhee Lee
POSTECH, lyhee@postech.ac.kr

Hyeseon Lee
POSTECH, hyelee@postech.ac.kr

Toward Optimal Churn Management: A Partial Least Square (PLS) Model

Hyeseon Lee
POSTECH
hyelee@postech.ac.kr

Yong Seog Kim
Utah State University
yong.kim@usu.edu

Yeonhee Lee
POSTECH
lyhee@postech.ac.kr

Hyunbo Cho
POSTECH
hcho@postech.ac.kr

ABSTRACT

In a very competitive mobile telecommunication industry business environment, marketing managers need a business intelligence model that allows them to maintain an optimal (at least a near optimal) level of churners very effectively and efficiently while minimizing the costs throughout their marketing programs. As a first step toward optimal churn management program for marketing managers, this paper focuses on building an accurate and concise predictive model for the purpose of churn prediction utilizing a Partial Least Square (PLS)-based methodology on highly correlated data sets among variables. A preliminary experiment demonstrates that the presented model provides more accurate performance than traditional prediction models and identifies key variables to better understand churning behaviors.

Keywords

Partial Least Square, Customer Relationship Management, Business Intelligence, Churn Management

1. INTRODUCTION

Although traditional mass marketing channels (e.g., advertisements in TV and newspapers) are still valid, these channels are not appropriate for customized marketing message. Therefore, the firms are searching for new ways of disseminating micromarketing message targeted toward a specific group of households that are most likely to open to the customized message. Thus, many models have been proposed to identify as many customers as possible who will respond to a specific solicitation campaign letter, or who will end further relationships with the firms. In particular, with exceptionally high annual churn rates (20–40%), the firms in mobile telecommunications industry tries to develop predictive models that accurately identify which customers are most likely to terminate the current relationship. Therefore, the optimal selection of customer targets (e.g., most probable churners) has been considered one of the most important factors for a successful customer relationship management (CRM) program.

Both marketing (Bult and Wansbeek, 1995; Rao and Steckel, 1995) and data mining researchers (Bhattacharyya, 2000; Chou et al., 2000) have presented various database marketing approaches for successful CRM programs. The simplest example is the RFM (recency, frequency, monetary) approach that targets households by using knowledge of the customer’s purchase history (Schmid and Weber, 1998; Gönül and Shi, 1998). In another study (Piatetsky-Shapiro and Masand, 1999), the profitability condition of a campaign was explicitly formulated as a function of the model performance along with campaign cost and revenue factors such as mailing costs and marginal revenue per identified positive record. A recent study (Kim, 2006) studies the effects of variable selection and class distribution on the performance of a primitive classifier system and a relatively more sophisticated classifier system in a customer relationship management (CRM) setting. Few models narrow their interests such as selecting prospects in the automotive industry (Gersten et al., 2000) and identifying most likely insurance buyers (Kim et al., 2005).

The goal of this paper is to propose a Partial Least Square (PLS)-based methodology that allows a marketing manager to maintain an optimal (at least a near optimal) level of churners very effectively and efficiently through her marketing programs. The detailed objectives are: (i) to build an accurate and concise predictive model for churner prediction based on PLS-based methodologies from the vast amount of data sets of highly correlated variables; and (ii) to conceptually design an effective churn control model after understanding key drivers of consumer behaviors.

In our approach, PLS is employed as the prediction modeling method because it places minimal demands on measurement scales, sample size, and residual distributions, and it is capable of handling a large number of highly correlated variables, measurement errors, and missing data (MacGregor et al., 1994). Further, PLS models naturally can be used for dimension reduction through variable selection mechanism based on the variable important in projection (VIP) scores. Therefore, it is possible that PLS models can be used not only for constructing highly accurate model but also for enhancing the comprehensibility of models by choosing a subset of the original predictive variables. By doing so, marketing managers can save a great amount of efforts and costs in identifying key determinants of churn behaviors of customers. However, eliminating many input variables may have different effects on the predictive accuracy of models depending on their representational powers and structural complexities. Therefore, this study aims to analyze the relationship between variable selection and the performance of several PLS models.

The remainder of this paper is organized as follows. Section 2 provides a review of PLS methodology. In Section 3, the research framework based on PLS models is introduced, and evaluation metrics to compare various predictive models on a data set collected by a mobile phone service provider are explained. In the following Section 4, experimental results of linear and nonlinear PLS models compared to Logit and random models will be presented. Finally, Section 5 provides the conclusion of the paper and suggests several direction of further research.

2. LITERATURE REVIEW

The PLS method is a multivariate projection approach to reduce the original large-scale data to lower dimensional data to deal with highly (both linearly and nonlinearly) correlated data between independent variables and dependent variables (Geldadi and Kowalski, 1986; Lakshminaraynan et al., 1997; Malthouse et al., 1997). One of main objectives in PLS analysis is to find a few PLS factors that explain most of the variation in both independent (=predictors) and dependent (=responses) variables. The PLS factors that explain most of the variation in responses using observed information of the predictors can consists of good predictive models for new responses. Therefore, the PLS method can be used as an alternative to well known models such as OSL regression, canonical correlation, or structural equation modeling (SEM) to investigate the relationship between predictors and response variables (http://faculty.chass.ncsu.edu/garson/PA765/pls.htm).

Note that the PLS method can model both multiple responses and multiple predictors variables, even when multicollinearity among predictors are suspected. Further, the PLS method is known to be robust when there are many observations with missing values in the data. However, it is difficult for the researcher to interpret loadings of the independent latent variables from the PLS method, and hence it is favored as a predictive technique and not as an interpretive technique. The PLS method can be implemented either as a regression model to predict response variables or as a path model to understand the structural relationship among records. In this study, the PLS method is used as a regression model to predict churners based on demographic, psychographic, and historical service usage information.

In order to describe the prediction models based on the PLS method, we first introduce some notations. Let X_j denote the j^{th} input variable ($j = 1,\ldots, m$) and let Y_k be the k^{th} dependent variable ($k = 1,\ldots, r$). Matrices $\mathbf{X} = (x_{ij})$ and $\mathbf{Y} = (y_{ik})$ are the historical data of predictor variables and the corresponding response variables, respectively, where x_{ij} is the i^{th} observation of X_j and y_{ik} is the i^{th} observation of Y_k ($i = 1,\ldots, n$). In the PLS method, $\mathbf{X}$ and $\mathbf{Y}$ are decomposed into a sum of series of lower dimensional matrices as follows:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \cdots + \mathbf{t}_A\mathbf{p}_A^T + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$$
$$\mathbf{Y} = \mathbf{u}_1\mathbf{q}_1^T + \mathbf{u}_2\mathbf{q}_2^T + \cdots + \mathbf{u}_A\mathbf{q}_A^T + \mathbf{F} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \tag{1}$$

where $\mathbf{T}(n\times A)$ and $\mathbf{U}(n\times A)$ represent the score matrix, while $\mathbf{P}(m\times A)$ and $\mathbf{Q}(r\times A)$ represent the loading matrix for $\mathbf{X}(n\times m)$ and $\mathbf{Y}(n\times r)$, respectively. To determine the dominant directions in which to project data, a maximal description of the covariance within **X** and **Y** is used as a criterion. The first set of loading matrices (the direction cosines of the dominant directions in the **X** and **Y**), $\mathbf{p}_1(m\times 1)$ and $\mathbf{q}_1(r\times 1)$, is obtained by maximizing the covariance between **X** and **Y**. The respective projection of the **X** and **Y** onto $\mathbf{p}_1$ and $\mathbf{q}_1$ gives the first set of score matrix, $\mathbf{t}_1(n\times 1)$ and $\mathbf{u}_1(n\times 1)$. This procedure is called "outer relation."

X and **Y** are also indirectly related through their scores by "inner relation," which is a functional mapping model from $\mathbf{t}_1$ to $\mathbf{u}_1$, that is, $\mathbf{u}_1 = h_1(\mathbf{t}_1)$. Denoting $\mathbf{E}_1 = \mathbf{X}$ and $\mathbf{F}_1 = \mathbf{Y}$, the residuals at this stage are computed by the deflation process:

$$\mathbf{E}_2 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^T = \mathbf{E}_1 - \mathbf{t}_1\mathbf{p}_1^T$$
$$\mathbf{F}_2 = \mathbf{Y} - \mathbf{u}_1\mathbf{q}_1^T = \mathbf{Y} - h_1(\mathbf{t}_1)\mathbf{q}_1^T = \mathbf{F}_1 - h_1(\mathbf{t}_1)\mathbf{q}_1^T \tag{2}$$

The procedure of determining the scores and loading matrices of the inner relation is continued (with the residuals obtained at each stage) until the required number of PLS dimensions (A) is extracted. In practice, the number of PLS dimensions is determined based on either the percentage of variance explained or the use of statistically sound approaches such as cross validation. The directions irrelevant in the data set (such as noises and redundancies) are confined to the error matrices, that is, **E** and **F**. Once P, Q and the function $\mathbf{H}=(\mathbf{h_1},\ldots,\mathbf{h_A})$ are estimated from the historical data, the quality data corresponding to a new process data set $\mathbf{X^o}$ can be predicted by the following model:

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} = \mathbf{H}(\mathbf{T})\mathbf{Q}^T + \mathbf{F} = \mathbf{H}(\mathbf{X}^o\mathbf{P})\mathbf{Q}^T + \mathbf{F} \tag{3}$$

Conceptually, the PLS method is based on decomposing the original variables of **X** and **Y** into the score variables of **T** and **U** that summarize a great deal of correlated and redundant information. Here, the column dimension of **T** and **U**, which is (A), is much less than that of $\mathbf{X}(m)$ or that of $\mathbf{Y}(r)$. It means that the higher dimension of the original variables is reduced to the lower dimension of the score variables. This is called the "dimension reduction".

In the nonlinear PLS models, the nonlinear functional relationship between **T** and **U**, $\mathbf{U} = \mathbf{H}(\mathbf{T})$, will be constructed by algorithms using neural networks or Gaussian kernel. Qin and McAvoy (1992) propose neural net PLS to approximate nonlinear mapping between the input and output score variables, and show benefit of a strategy that only a small size network is trained at one time. Malthouse et al. (1997) implement nonlinear PLS with feedforward neural network and apply to the model with multiple response variables. The nonlinear PLS based on neural network mapping requires quite computational time and it has limitation applying to large size data.

3. RESEARCH MODEL AND DATA SET

3.1 Research model

Our research framework consists of four sequential steps based on typical CRM processes, as illustrated in Figure 1. As the first step, it is necessary to preprocess raw data into a readily available format for further analysis. In this study, two different techniques—eliminating records with missing values and variable selection—are used separately or together for preprocessing raw data. Once preprocessed data sets from raw data are obtained, three different types of classifiers (Logit regression, linear and nonlinear PLS models) are calibrated and evaluated. In this process, the performance of a random model will be also evaluated to highlight additional gains of predictive power by using any one of three intelligent models. As one of intelligent model types, several linear

PLS models will be calibrated and evaluated in terms of comprehensibility, computational complexity, and predictive power related metrics such as hit ratio. Then nonlinear PLS model and Logit regression models are constructed and compared with the most accurate linear PLS model. Note that multiple Logit regression models with different pre-defined significance values to add new variables into the final predictive models will be considered to estimate the effectiveness of dimension reduction via variable selection with forward variable selection and the PLS method. We also intend to compare linear PLS model to nonlinear model to see if a nonlinear relationship mapping between predictors and response variables can boost predictive power of the linear PLS models. Finally, managerial insights extracted from experimental results are discussed to help data analysts and marketing managers develop a successful churn management program by improving service quality and developing management strategies based on hardware replacement, complaint management, and service quality improvement.

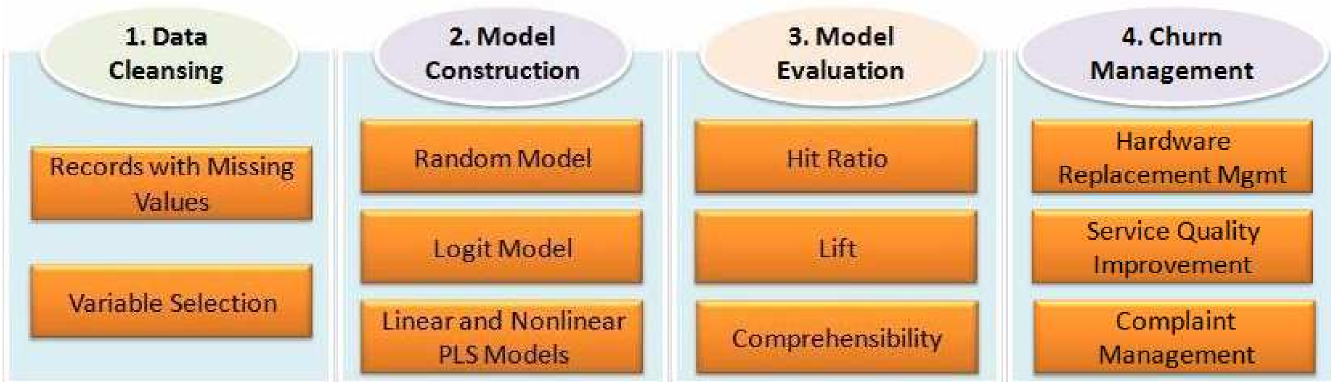

Figure 1 Research framework

3.2 Data set

The data sets used in this study are provided by the Teradata Center for CRM at Duke University, and the original data set for calibration has 171 predictor variables of 100,000 observations. The complete set of variables includes three types of variables: behavioral information such as minutes of use, revenue, handset equipment; company interaction information such as customer calls into the customer service center; and customer household demographics. For each customer, churn was calculated based on whether the customer left the company during the 31- to 60-day period after the customer was originally sampled. Although the actual percentage of customers who left the company in a given month is approximately 1.8%, churners in the original data set were oversampled to create roughly a 50–50 split between churners and nonchurners. However, the test data set with 51,306 observations are expected to represent a realistic churning rate, 1.8%.

As a preprocessing step for further analysis, the original data sets are preprocessed as follows. First, most categorical variables are excluded because of high missing rate or being encoded into multiple binary variables which makes low predictive power. We include only 11 categorical variables which are either indicator variables or countable variables such as number of handsets and number of subscribers. This is because each categorical variable has very little predictive power in general (Rossi et al., 1996). Second, continuous variables with more than 20% of missing values are eliminated. We take 123 predictors including 11 categorical variables and 112 continuous variables in data preprocessing step. Finally, records with missing values in the data set with 123 predictors are removed from further analysis. After preprocessing steps, the training set contains 67,181 observations with 32,862 churners, while the test set contains 34,986 observations with 619 churners, respectively.

3.3 Evaluation metrics

In this study, hit rate and lift trend curve are used to numerically quantify the predictive power of models and graphically represent the performance for easy comparison, respectively. The hit rate is defined as the number of correctly identified churners out of churner candidates in this study. When only *x%* of customers predicted most likely to churn are considered for the model evaluation, it is called hit rate at target point *x%*. For example, if the model is required to select 1000 customers who are most likely to churn from 10,000 observations, and 200 of them turn out to be actual churners, then a hit rate at target point 10% (1000/10,000=10%) is 20% (200/1000=20%). The lift trend curve shows a lift at a target point *x%*, where a lift is a ratio of the hit rate of a predictive model divided by the hit rate of a random model. This paper uses the raw number of correctly identified churners over small target points due to the limited budget and time constraints to develop marketing programs.

4. EXPERIMENTAL RESULTS

4.1 Forward variable selection vs. PLS variable selection

In order to measure the effectiveness of dimension reduction of the proposed PLS models, a very simple variable selection procedure, forward variable selection, is performed. Once a set of informative variables is selected through the forward selection, it is used as predictor variables of logit regression model to estimate the likelihood of becoming churners. Note that the forward variable selection procedure starts with the empty set of variables and greedily adds variables that most improve performance based on χ^2 scores. It stops adding new variables when there is no additional variable that satisfies the predefined significance level (e.g., α=0.05) for entry into the model. Two different values of significance levels are used (α=0.05 and α=0.15). From the perspective of model interpretation, a predictive model with a parsimonious variable subset (i.e., α is set to 0.05) is preferred as long as the performance of two models is compatible because of improved comprehensibility. Further, the choice of a simpler model is consistent with the well known “Occam’s razor” principle, which states that if all other things being equal, a simpler model generalizes better and hence preferred (Blumer et al., 1987).

The PLS models can also be used to reduce data dimension using the information of variable importance in projection (VIP) scores of each variable. According to Wold (1994), the researcher may safely remove any independent variable which has a small VIP (< .8) and a small absolute value of regression coefficient. As an exploratory study, we use a set of VIP cut-off criteria values (0, 1.0, 1.2, and 1.5). For example, $PLS^{1.0}$ model is a PLS model with all variables whose VIP scores is greater than or equal to 1.0, while PLS^{all} model utilizes all variables (i.e., PLS^{0}). Therefore, $PLS^{1.5}$ model is the most parsimonious PLS model and PLS^{all} model is the most comprehensive model. Due to the limited space, we only show 10 variables in Table 1 selected by two models: the forward variable selection with α=0.05 and variable selection based on $PLS^{1.0}$ model. These variables are listed in the selection order in the process of forward variable selection or VIP scores in the PLS variable selection, and variables in bold are variables selected by both models.

Table 1 List of variables selected via forward variable selection and PLS

Variable selection method	Variable subsets
Forward variable selection (α=0.05)	1. **Number of days of current equipment**, 2. **Handset (refurbished or new)**, 3. **Range of overage revenue**, 4. Mean number of monthly minutes of use, 5. **Average monthly minutes of use over the life of the customer**, 6. **Account spending limit**, 7. **Total number of months in service**, 8. Range of number of minutes of use, 9. Range of total monthly recurring charge, 10. Number of unique subscribers in the household

PLS variable selection (VIP>=1.0)	1. **Number of days of current equipment**, 2. **Handset (refurbished or new)**, 3. **Account spending limit**, 4. **Average monthly minutes of use over the life of the customer**, 5. **Range of overage revenue**, 6. Range of revenue of voice overage, 7. **Total number of months in service**, 8. Percentage change in monthly minutes of use vs previous 3 month average, 9. Range of overage minutes of use, 10. Average monthly number of calls over the life of the customer

We first notice that most of selected variables reflect customer's usage behaviors and interactions with the company, which are in line with marketing science work (Rossi et al., 1996). One variable, Number of days (age) of current equipment, is selected as the most importance in churn prediction by both models. It is also noted that a variable indicating that a subscriber's phone was refurbished or new one (Handset-refurbished or new) is recognized as the second most important variable by the two models. Further, six out of top 10 listed variables are selected by the two methods.

The more interesting fact is that the first two most predictive variables are variables related to mobile handset and it is not very difficult to link these variables to churn behavior. For example, many mobile phone service subscribers in USA are required to enroll into a two-year contract on the condition that certain handsets are given free. Most subscribers are not likely churn to another service provider until the contract expires to avoid financial penalty. Only when the contract period ends and new handset models with more functions and better interfaces are available from other service providers, the user may switch to a new service provider. Therefore, the number of days of current equipment is a strong indicator when the current subscriber may switch to another service provider or stay with the current service provider. The second most important variable, Handset-refurbished or new, implies the almost the same information. Typically, if a customer purchases the new handset, she is most likely to purchase as a promotion package from the current service provider and hence stay with the current subscriber until the end of a contract period. However, if she purchases a used or refurbished handset without any obligation from the current service provider, she can either stay with the current service provider or switch to a new service provider.

From marketing managers' perspective, this finding implies that marketers must pay attention to their current users who has used the current handset for almost their contract periods or whose contract period ends in near future. Once users with higher probability of churning are identified, marketing managers immediately contact them to offer customized micromarketing messages before the contract period ends or users switch to another service provider. For example, about one month before the end of the contract period, the marketer may send an appreciation letter to offer a new handset with more advanced functions at a reasonable price as long as the current user agrees to stay with the current service provider for another contract period. Other variables of user's service usage patterns are informative to predict churn behaviors of the users. For example, heavy service users with higher range of overage revenue and number of minutes of use, or higher value of average monthly number of calls over the life of the customer are most likely to stay with the current service provider as long as they are satisfied with the current service quality. In addition, when users do not use the mobile service as much as they have been in terms of minutes of use with the current service provider, it is very likely that they might switch to another service provider. Finally, the number of unique subscribers in the household and account spending limit variable may affect the user's churning behavior negatively.

4.2 Churner identification with linear PLS Models

In this section, we present the predictive performance of linear PLS models in a prediction task to correctly identify churners. Figure 2 presents hit rates of linear PLS models combined with four different cut-off values of VIP scores to reduce data dimension. Unfortunately, Figure 2 shows that PLS models with more input variables are more predictive and should be used for predictive purposes. Note, however, that all PLS models perform significantly better than a random model of which proportion of hit records is always equal to the proportion of chosen records. Both PLS^{all} and $PLS^{1.0}$ models identify more than 30% of churners correctly at a target point 20%. In particular, the performance of $PLS^{1.0}$ model is very compatible to PLS^{all} model when a small portion of records are chosen for churn management program, and hence could be used as an alternative to PLS^{all} model when both costs and benefits of churn management program are considered. Note that with more targets for a churn management program, a market manager needs to spend more money and time to distribute campaign messages and offer a discounted service fees or free handsets.

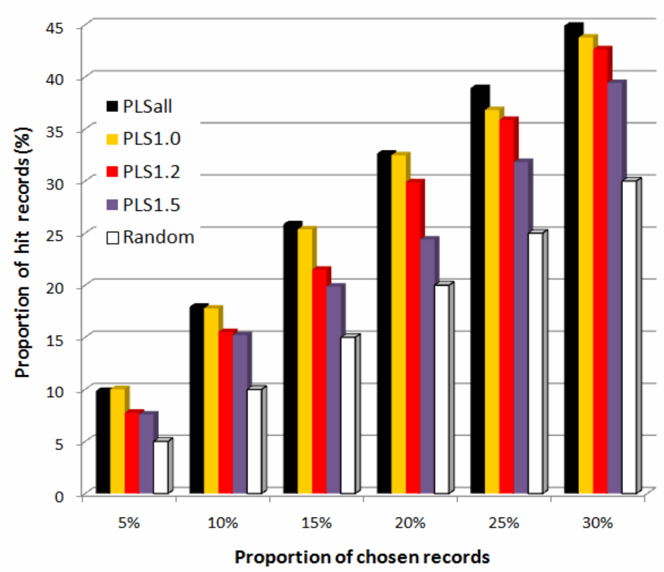

Figure 2 Hit rates of linear PLS models

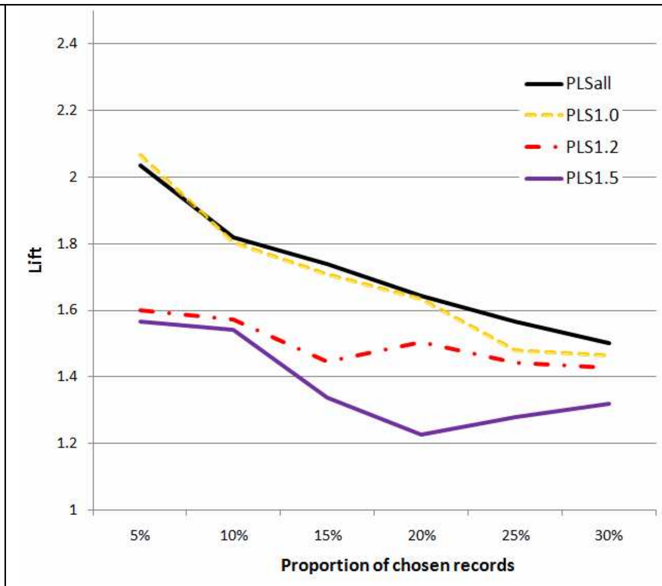

Figure 3 Lift trends of linear PLS models

Figure 3 presents the relative performance of all linear PLS models compared to a random model. While *x* axis still represents a set of proportion of chosen records, *y* axis represents a set of lifts defined as the hit rate of model A divided by the hit rate of a random model at a chosen record proportion. Therefore, lift values that are higher than 1.0 indicate that the compared model performs better than a random model. We immediately note that both PLS^{all} and PLS$^{1.0}$ models identify twice as many hit records (=correctly identified churners) as a random model, while other PLS models such as PLS$^{1.2}$ and PLS$^{1.5}$ also identify about 60% more churners at a selection point 5%. Naturally, all PLS models show the decreasing trends of lift values as more users with lower estimated probability of churning are considered in a pool of churner candidates. However, we note that parsimonious PLS models with higher VIP cut-off scores, PLS$^{1.2}$ and PLS$^{1.5}$, show relatively stable lift trends compared to PLS models with lower VIP cut-off score such as PLS^{all} and PLS$^{1.0}$ model partially because of their relatively lower lift values. However, we also attribute this finding to the fact that parsimonious PLS models typically generalize better with a smaller number of input variables for a prediction task.

4.3 Churner identification with nonlinear PLS model

We also compare the predictive performance of a nonlinear PLS model and another very popular model in marketing community, logit models. We also included one of linear PLS model, PLS^{all}, that shows the best performance out of all linear PLS models, and a random model for comparison purposes. We use the same hit rate and corresponding lift trend curve for comparing the predictive performance of predictive models.

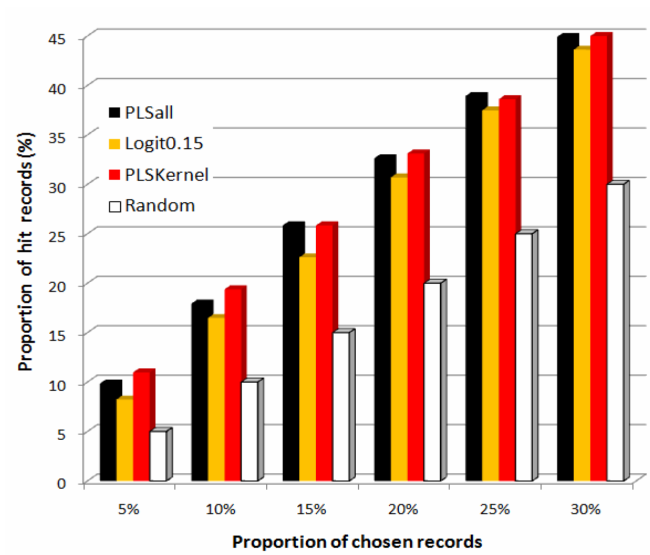	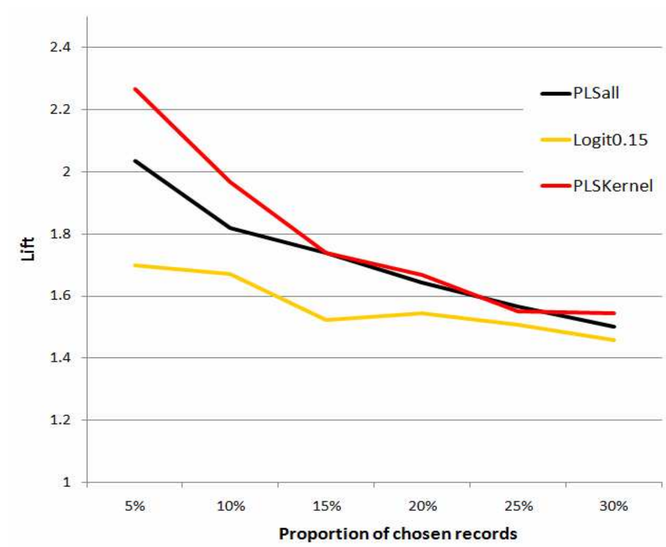
Figure 4 Hit rates of nonlinear PLS and logit models	Figure 5 Lift trends of nonlinear PLS and logit models

Two Logit models are implemented, Logit$^{0.15}$ with α=0.15 and Logit$^{0.05}$ with α=0.05, and Logit$^{0.05}$ is a more parsimonious model because the forward variable selection method stops adding new variables with a stricter significance level. Although two models are very compatible, Logit$^{0.15}$ model was slightly better in terms of predictive power and hence was chosen for predictive performance comparison. Note, however, that we prefer and recommend Logit$^{0.05}$ model to marketing managers because it helps marketing managers to better understand churn behaviors from fewer variables and hence develop new micromarketing programs easily. In order to implement nonlinear PLS model, PLS^{kernel}, we use a Gaussian kernel function, $K(x_i - x_j) = exp(-\|x_i - x_j\|^2 / k)$, where x_i and x_j represent i^{th} and j^{th} vector observation of X, and k is a parameter value of Gaussian kernel function. We compute the values of the kernel function from 6,000 samples with even class distribution from training data to lessen computational overload after we standardized the values of observations for each variable, and the value of k is subjectively set to 200. Figure 4 and 5 present hit rates and hit rate trends of four predictive models, respectively: linear PLS^{all}, Logit$^{0.15}$, PLS^{kernel}, and a random model. Both PLS^{all} and PLS^{kernel} models perform well. In particular, PLS^{kernel} relatively performs better at 5% and 10% target points, returning higher lift values compared to the other models.

6. CONCLUSION AND FUTURE RESEARCH

In this paper, we introduce and build an accurate and parsimonious model based on PLS method to predict churn behaviors on highly correlated data sets among variables on which regular regression models cannot be built. In terms of quality of the chosen subset of features, a subset of features chosen via the PLS model is very similar to those selected by popular forward feature selection algorithm, and share many features in common. In terms of predictive accuracy of the PLS model measured by hit rates and shown in lift trend curves, all PLS models perform well and PLS^{all} is the most predictive model. Note, however, that while PLS^{all} model performs slightly better than PLS$^{1.0}$ model, PLS$^{1.0}$ model is much more parsimonious than PLS^{all} and hence may be preferred by marketing managers.

A promising direction of future research is to extend the current research framework to develop an optimal churn management system. In the proposed churn management system, the chosen subset of features from the PLS predictive model are carefully investigated and divided into controllable and uncontrollable variables. Then, a mathematical formulation will be applied to minimize the churn rate (or maximize financial benefits from churn management program) by changing the values of controllable variables with constraints that costs involved in manipulating controllable variables should be less than or equal to the overall budget limit.

ACKNOWLEDGEMENTS

The author wishes to thank the Teradata Center for CRM at Duke University for making the data sets available.

REFERENCES

1. Abrahamsson, P., Warsta, J., Siponen, M., and Ronkainen, J. "New Directions on Agile Methods: A Comparative Analysis," *Proceedings of the 25th International Conference on Software Engineering*, 2003, pp. 244–254.
2. S. Bhattacharyya, "Evolutionary algorithms in data mining: Multi-objective performance modeling for direct marketing," *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 465–473.
3. A. Blumer, A. Ehrenfeucht, D. Haussler, and M.K. Warmuth, "Occam's razor," *Information Processing Letters* (24), 1987, pp. 377–380.
4. J.R. Bult, and T. Wansbeek, "Optimal selection for direct mail," *Marketing Science* (14:4), 1995, pp. 378–394.
5. P.B. Chou, E. Grossman, D. Gunopulos, and P. Kamesam, "Identifying prospective customers," *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 447–456.
6. Y. Kim, "Toward a successful CRM: Variable selection, sampling, and ensemble," *Decision Support Systems* (41:2), 2006, pp. 542–553.
7. Y. Kim, W. N. Street, G. J. Russell, and F. Menczer, "Customer targeting: A neural network approach guided by genetic algorithms," *Management Science* (51:2), 2005, pp. 264–276.
8. P. Geldadi and B. Kowalski, "Partial least-squares regression: A tutorial," *Analytica Chemica Acta* (185), 1986, pp. 1–17.
9. F. Gönül, and M. Z. Shi, "Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models," *Management Science* (44:9), 1998, pp. 1249–1262.
10. W. Gersten, R. Wirth, and D. Arndt, "Predictive modeling in automotive direct marketing: tools, experiences and open issues," *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 398–406.
11. S. Lakshminaraynan, S. Shah, and K. Nandakumar, "Modeling and control of multivariable processes: Dynamic PLS approach," *AIChE Journal*, (43:9), 1997, pp. 2307–1322.
12. J.F. MacGregor, C. Jaeckle, C. Kiparissides and M. Koutoudi, "Process monitoring and diagnosis by PLS methods," *AIChE Journal* **(**40:5)**,** 1994, pp. 826–836.
13. E. Malthouse, A. Tamhane, and R. Mah, "Nonlinear partial least squares," *Computers and Chemical Engineering* (21:8), 1997, pp. 875-890.
14. G. Piatetsky-Shapiro and B. Masand, "Estimating campaign benefits and modeling lift," *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining*, 1999, pp. 185–193.
15. V.R. Rao and J. H. Steckel, "Selecting, evaluating and updating prospects in direct mail marketing," *Journal of Direct Marketing* (9:2), 1995, pp. 20–31.
16. P.E. Rossi, R. McCulloch, and G. Allenby, "The value of household information in target marketing," *Marketing Science* (15:3), 1996, pp. 321–340.
17. S.J. Qin and T.J. McAvoy, "Nonlinear PLS modeling using neural networks," *Computers and Chemical Engineering* (16: 4), 1992, pp. 379-391.
18. J. Schmid and A. Weber, *Desktop Database Marketing*, NTC Business Books, Lincolnwood, IL, 1998.

19. K. Song, P. Y. Jang, H. Cho, and C-H Jun. "Partial least square-based model predictive control for large-scale manufacturing processes," *IIE Transactions* (34), 2002, pp. 881-890.
20. S. Wold, "PLS for multivariate linear modeling," In: H. van de Waterbeemd, Editor, *QSAR: Chemometric Methods in Molecular Design. Methods and Principles in Medicinal Chemistry*, Verlag–Chemie, Weinheim, Germany, 1994.