December 2004

Applying Data Mining to Telecom Churn Management

Shin-Yuan Hung
National Chung-Cheng University

Hsiu-Yu Wang
National Chung-Cheng University

Follow this and additional works at: http://aisel.aisnet.org/pacis2004

Applying Data Mining to Telecom Churn Management

Shin-Yuan Hung
Department of Information Management
National Chung-Cheng University
Taiwan, ROC
syhung@mis.ccu.edu.tw

Hsiu-Yu Wang
Department of Information Management
National Chung-Cheng University
Taiwan, ROC
HsiuYu.Wang@msa.hinet.net

Abstract

Taiwan deregulated its wireless telecommunication services in 1997. Fierce competition followed, and churn management becomes a major focus of mobile operators to retain subscribers via satisfying their needs under resource constraints. One of the challenges is churner prediction. Through empirical evaluation, this study compares various data mining techniques that can assign a "propensity-to-churn" score periodically to each and every subscribers of a mobile operator. The results indicate that both decision tree and neural network techniques can deliver accurate churn prediction models by using customer demographic, billing information, contract/service status, call detail records, and service change log.

Keywords: Churn Management, Wireless Telecommunication, Data Mining, Decision Tree, Neural Network

1. Introduction

Taiwan opened its wireless telecommunication services market in 1997, with license granted to six mobile operators. Competition has been fierce from the beginning. For acquisition, operators have significant network investment to provide ubiquitous access and quality communications. The market saturated in five years, and merges and acquisitions have reduced the number of mobile operators from six to four by the end of 2003.

When the market is saturated, the pool of "available customers" is limited and an operator has to shift from its acquisition strategy to retention because the cost of acquisition is typically five times higher than retention. As Rob Mattersion (Mattersion 2001) noted, "For many telecom executives, figuring out how to deal with Churn is turning out to be the key to very survival of their organizations."

Based on marketing research (Berson et al. 2000), the average churn of a wireless operator is about 2% per month. That is, a carrier lost about a quarter of its customer base each year. Furthermore, Figure 1 suggests that Asian telecom providers face a more challenging customer churn than those in the other parts of the world.

From business intelligence perspective, churn management process under the CRM (Customer Relationship Management) framework consists of two major analytical modeling efforts: predicting those who are about to churn and assessing the most effective way that an operator can do (including "do nothing") in terms of retention. This research focuses on the former. It intends to illustrate how to apply IT technology to facilitate telecom churn management. Specifically, this research uses data mining techniques to find a best model of predictive churn from data warehouse to prevent the customers turnover, further to enhance the competitive edge.

This paper shares the result of the research. Section 2 defines some basic concepts (and rationale) that we use in the research, Section 3 describes our research methodology, and Section 4 presents the findings. Section 5 concludes our presentation.
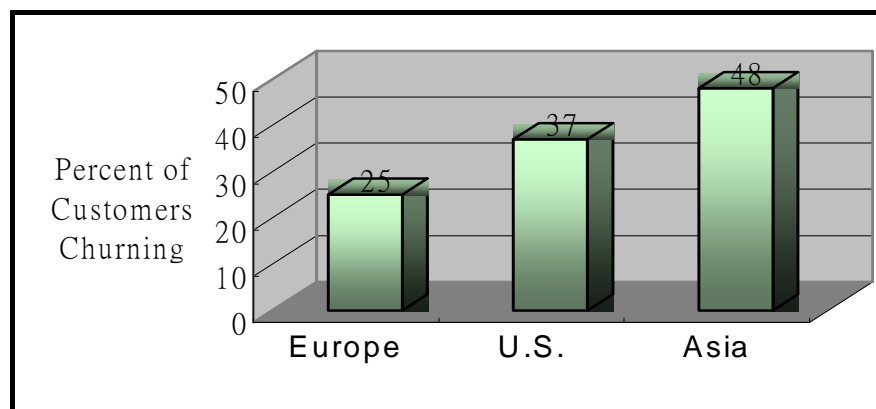

Figure 1 Annual Telecom Operator Customer Churn Rate by Region (Mattersion 2001)

2. Basic Concept

2.1. Churn Management

Berson et al. (Berson et al. 2000) noted that "Customer Churn" is a term used in the wireless telecom service industry to denote the customer movement from one provider to another, and "Churn Management" is a term that describes an operator's process to retain profitable customers. Similarly, Kentritas (Kentritas 2001) thought that the term Churn Management in the telecom services industry is used to describe the procedure of securing the most important customers for a company. In essence, proper telecom operator customer management presumes an ability to forecast the customer decision to move from one service provider to another, a measurement of "customer profitability," and a set of strategic and tactic retention measures to reduce the movement.

In practice, an operator can segment its customers by "profitability" and focus retention management only on those profitable segments, or score the entire customer base with "propensity to churn" and prioritize the retention effort based on profitability and churn propensity. However, the telecom services industry is yet to standardize a set of "profitability" measurements (e.g., current versus life-time, business unit versus corporate, account versus customer, "loyalty" versus "profitability," etc.) This research focuses on churn prediction, and assumes low correlation between "profitability" and "propensity to churn" to simplify the modeling framework.

2.2. Data Mining

Kurt Thearling (Thearling 1999) proposed that Data Mining is "the extraction of hidden predictive information from large databases," a cutting-edge technology with great potential to help companies dig out the most important trends in their huge database. Emerging data mining tools can answer business questions that have been traditionally too time-consuming to solve. Miguel Lejeune (Lejeune 2001) addressed that Data Mining Techniques allow the transformation of raw data into business knowledge. The SAS Institute (SAS 2000) defines data mining as "the process of selecting, exploring and modeling large amount of data to uncover previously unknown data patterns for business advantage". Consequently, we would say that data mining is applying data analysis and discovery algorithms to detect patterns over data for prediction and description.

With sufficient database size and quality, data mining technology can provide business intelligence to generate new opportunities. In the CRM space, data mining techniques most commonly used include clustering, associations, rule induction, genetic algorithm, decision tree, and neural network.

2.3. Data Mining Application
Table 1 below summarizes some data mining functionalities, techniques, and applications in the CRM space.

Table 1: Data Mining Functionalities, Techniques, and CRM Applications

Functionality	Technique	Application
Association	• Set Theory • Statistics • Bayesian Classification	• Cross Sell
Estimation	• Neural network • Statistics • Time Series	• Exchange Rate Estimation • Stock Price Estimation
Classification	• Decision Tree • Fuzzy • Neural Network • Genetic Algorithm	• Credit Embezzle • Market Segmentation
Prediction	• Regression • Neural Network • Decision Tree	• Churn Prediction • Fraudster Prediction
Segmentation	• Neural Network • Statistics • Genetic Algorithm • Decision Tree	• Market Segmentation

Source: Collected by this Study

3. Churn Prediction Data Mining Assessment Methodology
The purpose of this research is to assess the performance of various data mining techniques when applied to churn prediction. The methodology consists of three parts:

(1). An IT infrastructure to facilitate our research, which includes a common customer base, attributes and transactions, modeling parameters, model results, etc.,

(2). A model-independent knowledge discovery process to discover customer behavior prior to churn, by using data mining techniques, and

(3). A set of measurements to quantify the performance of models developed by different modeling tools, such as decision tree and neural network.

3.1. Churn Management Research Framework
Figure 2 shows the conceptual infrastructure that we use to assess the performance of various churn prediction models built via data mining techniques. We built the infrastructure on a Data Warehouse with tables, views, and micros to facilitate the following model development and assessment processes.

(1). Identify data items of interest from customer behavior attributes that can differentiate between churners and non-churners,

(2). Extract, transform, and derive variables from identified data items,

This research selected Decision Tree, Neural Network and K-means cluster as data mining techniques to build predictive models or segment customers.

Note that in addition to conduct empirical researches, we can use the same IT infrastructure to collect, analyze, detect, and eliminate major customer churn factors. This "closed loop" infrastructure is imperative to business management as we manage churn to sustain our relationship with customers.
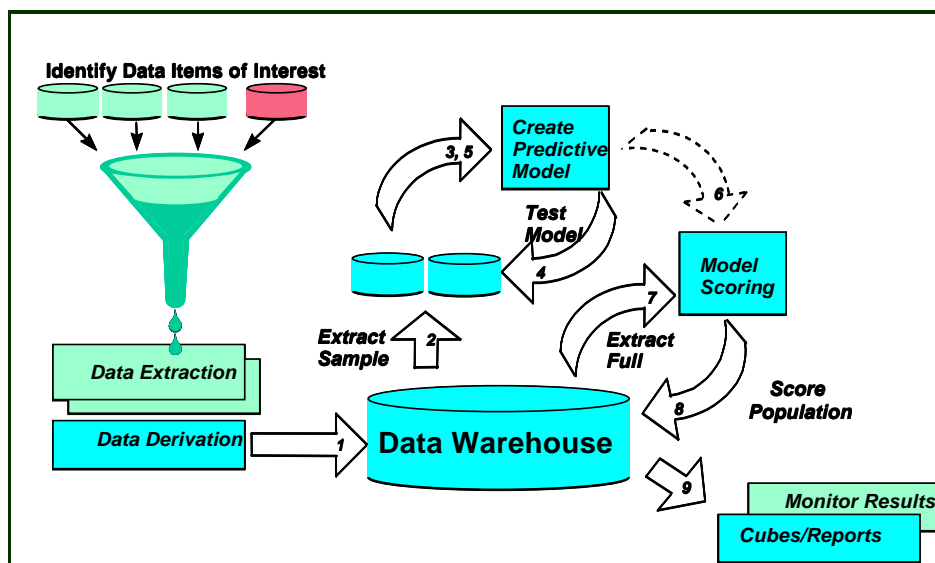

Figure 2: IT Infrastructure of Model Assessment Process (source: NCR Inc.)

3.2. Prediction Model Creation Process
Figure 3 shows our process of creating a predictive model.

I. Define Scope
   In this study, we focus on the post-paid subscribers who pay monthly fee and were activated for at least 3 months prior to July 1, 2001. Churner is defined as a subscriber who is voluntary to leave; non-churner is the subscriber who is still using this operator's service. Moreover, we used the latest six months' transactions of each subscriber to predict customer's churn probability of the following month. The transaction data include billing data, call detail records (CDR), customer care, etc.

II. Exploratory Data Analysis (EDA)
   The purpose of EDA is to explore from customer database those possible variables that can characterize or differentiate the customer behavior. For the variables extraction, we interviewed telecom experts, such like telecom business consultants, marketing analysts, customer carer, and sales of mobile provider to identify churn causes or symptoms prior to customer churn, such as "contract expired," "low usage," or "query about terminating the contract."
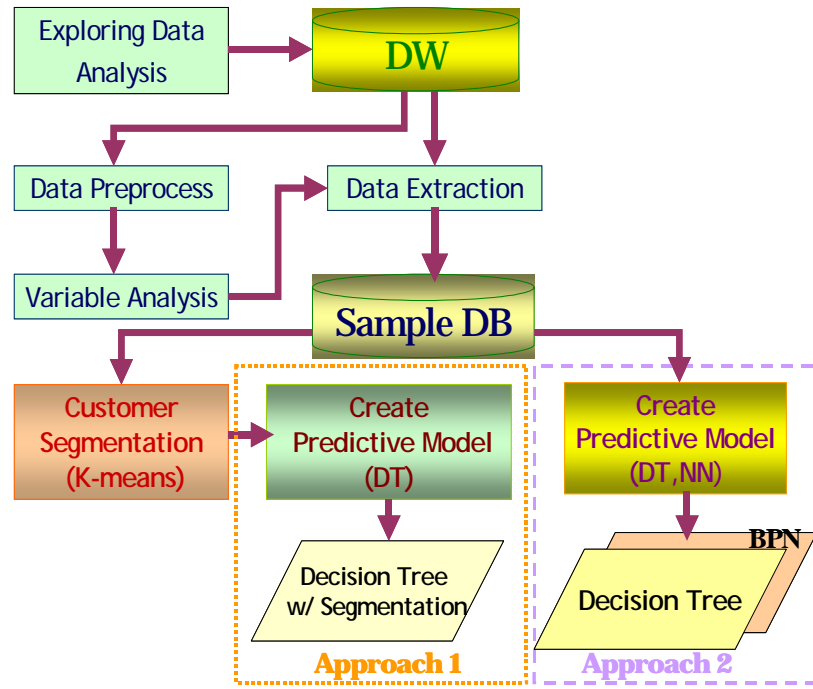
Figure 3. Process of Predictive Model Creation

III. Data Preprocessing, Variable Analysis and Selection, and Data Extraction

Based on the results of interviewing with experts, we extract some possible variables from the customer database as an analytical base of EDA to determine which variables are useful to differentiate between churners and non-churners.

For each of the causes/symptoms gained from interviews, we start to determine if we can observe similar customer behavior from the database. For example, to the symptom "contract expired" we can define a variable "Number of Days between Today and Contract Expiration Date" to test its correlation with customer churn, where "Today" is the date of prediction. Depending on the variable type, we can use different statistical test tools, such as Z-test. (That is, we examine variable significance by Z-score, 99%, and selected if its Z-score is over 3.)

Note that "Contract Expiration Date" must be a quality data field ("table column") in the database. Otherwise, the statistical inference based on this variable would be invalid. A significant effort in Data Preprocessing is to resolve data quality issues related to unspecified business rules or business rules not enforced in the business process.

Note also that we can have alternative variable definitions, such as "1 if Today is later than Contract Expiration Date and 0 otherwise." It is an iterative process to define the variables, identify the table columns, specify the calculation formula, test the validity of statistical inference, and select useful variables for modeling. Data Extraction is a formalized system integration process to ensure data quality and code optimization in modeling, production (e.g., scoring), and model maintenance.

IV. Machine Learning (Model Creation)

We took two approaches to assess how models built by Decision Tree (C5.0) and Back Propagation Neural Network (BPN) techniques perform.

In Approach 1, we used K-means clustering methods to segment the customers into 5 clusters according to their billing amount (to approximate "Customer Value"), tenure months (to approximate "Customer Loyalty"), and payment behaviors (to approximate "Customer Credit Risks"). Then we create a Decision Tree model in each cluster (see Approach 1 in Figure 3). This is to assess if the churn behaviors are different in different "Value-Loyalty" segments.

In Approach 2 (see approach 2 in Figure 3), we used Neural Network to segment customers, followed by Decision Tree modeling. This is a technology assessment to test if BPN can improve DT prediction accuracy.

3.3. Model Performance Evaluation

The real world wants to know how accurate a model predicts and how long before the model requires maintenance.

To assess the model performance, we use LIFT and Hit Ratio. The following chart illustrates their definitions, where A is the number of subscribers predicted to churn in the predictive time window and actually churned, B is the number of subscribers predicted to churn but not churned, etc.

Hit Ratio is defined as A/(A+B), instead of (A+D)/(A+B+C+D). This is a model effectiveness measurement in predicting churners instead of predicting all customer behavior in the predictive window.

To assess LIFT, we have to rank order all customers by their churn score, and define Hit Ratio (X%) as the Hit Ratio of the "X% Customers with Top Churn Score." LIFT(X%) is then defined as the ratio of Hit Ratio (X%) to the overall monthly churn rate. For example, if the overall monthly churn rate = 2%, X=5, and Hit Ratio (5%) = 20%, then LIFT (5%) = 20%/2% = 10. LIFT is a measure of productivity with modeling: With random sampling of the entire customer base you would get 2% churners, focusing instead on the 5% Top-Churn-Score customers you would get 20% churners. (Note that in this case, the top 5% contains 50% of the total churners.)

Capture Ratio (X%) = A/(A +C)
Hit Ratio = A/(A +B)

	Prediction	
	1	0
Actual 1	A	C
Actual 0	B	D
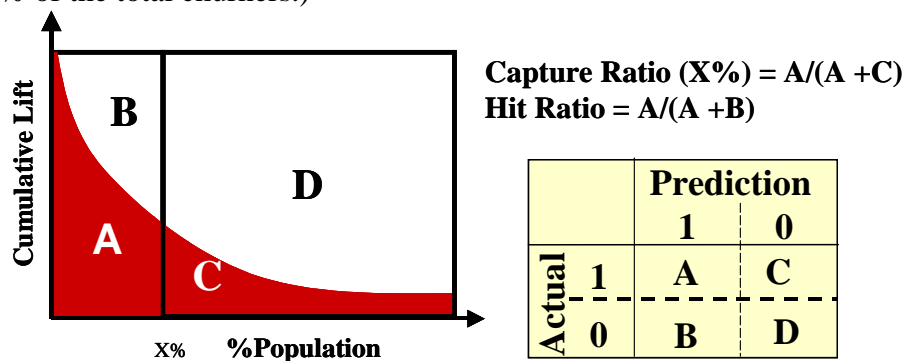
Cumulative Lift (X%)
= Churn Rate of subscribers within X% Population / Population Churn Rate
= Hit Ratio (X%) / Overall Churn Rate

To assess model robustness, we monitor each month's model Hit Ratio and LIFT for extended period of time to detect degradation.

4. Empirical Finding

4.1. Data Source

A wireless telecom company in Taiwan provides their customer related data. To protect customer privacy, the data source includes data of about 160,000 subscribers, including 14,000 churners, from July 2001 to June 2002, randomly selected based on their telephone numbers.

4.2. Exploring Data Analysis Results

We got possible variables from other researches and telecom experts' interviews. We then analyzed these variables with z-test from four dimensions and listed significant variables of churn below, based on our analysis database.

- Customer Demography
  - Age: Analysis shows that the customers between 45 and 48 have a higher propensity to churn than population's churn rate.
  - Tenure: Customers with 25 to 30 months' tenure have high propensity to churn. A possible cause is that most subscription plans have a 2-year contract period.
  - Gender: Corporate accounts' churn probability is higher than others. A possible cause is that when employees quit, they lose corporate subsidy in mobile services.

- Bill and Payment Analysis
  - Monthly Fee: The churn probability is higher of customers with a monthly fee less than $100 NT or between $520 and $550.
  - Billing Amount: The churn probability tends to be higher of customers whose average billing amount over six months is less or equal to $190 NT.
  - Count of Overdue Payment: The churn probability is higher for customers with less than 4 counts of overdue payments in the past 6 months. In Taiwan, if the payment is two months' overdue, the mobile operator will most likely suspend the mobile service until fully paid. This may cause customer dissatisfaction and churn.

- Call Detail Records Analysis
  - In-Net Call Duration: Customers who don't often make phone calls to others in the same operator's mobile network are more likely to churn. In-Net unit price is relative lower than that of other call types. Price-sensitive subscribers may leave for other mobile operator that his/her friends subscribe.
  - Call Type: Customers who often make PSTN or IDD calls are more likely to churn than those who make more mobile calls.

- Customer Care/Service Analysis
  - MSISDN Change Count: Customers who have ever changed phone number or made twice or more changes in account information are more likely to churn.
  - Count of Bar & Suspend: Customers who have ever been bared or suspended are more likely to churn. In general, a subscriber will be bared or suspended by the mobile operators due to payment overdue.

Table 2 summarizes variables significant to differentiate between churners and non-churners from EDA. We use those variables for machine learning.

Table 2. Significant Variables of Churn

Dimension	Items
Demography	Gender, Age, Area, Tenure
CDR	Inbound Call, Outbound Call, Demestric Call,
Bill/Payment	Bill Amout, Payment, Overdue Paymemt, Monthly Fee
Custer Service	Inquire, Phone No Changed, Bar/Suspend

4.3. Customer Segmentation

To segment customers by loyalty, contribution, and usage, we selected Bill Amount, Tenure, MOU (Outbound Call Usage), MTU (Inbound Call Usage), and Payment Rate as variables and used K-Means to model the customers into 5 clusters. To generate roughly the same number of subscribers in each of the 5 clusters, we divided the customers equally into three segments High, Medium, and Low for each variable.

Table 3 and Figure 4 summarize the clustering results. Note that each cluster has its unique characteristics:

Table 3. Customer Segmentation---Cluster

Cluster ID	Tenure	Bill AMT	MOU	MTU	PYMT Rate	% of Population	ChurnRate
C1	H	H	H	H	M	32.9%	0.50%
C2	L	L	L	L	L	26.8%	1.19%
C3	H	M	M	M	L	14.0%	0.32%
C4	L	M	M	M	L	16.7%	0.30%
C5	M	M	M	M	H	9.6%	1.37%

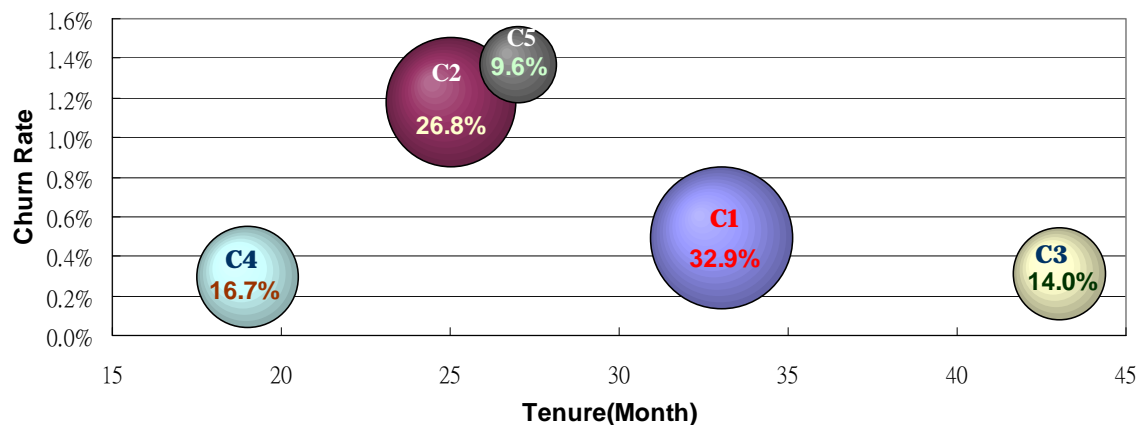Note: L: Low, M: Medium, H: High

Figure 4. Cluster Distribution (churn Rate vs. Tenure)

4.4. Supervise Machine Learning
4.4.1. Decision Tree

In this research we use Decision Tree technique to create many models under different scenarios. One of the scenarios is to create a Decision Tree model for all the customers. Another one is to create separate Decision Tree models in each of the customer segments.

1. Decision Tree Modeling Without Customer Segmentation
    From EDA we select about 40 variables for C5.0 Decision Tree modeling. We use many different training sets to create models, and use the same data set for model validation and performance test. We compare model performance based on their LIFT(10%), that is, the LIFT of the 10% subscribers with the top churn scores.

    The churn rate of whole population is only 0.71%. We over-sample the churners to get a

higher churn rate training set for machine learning. Table 4 summarizes the results of six different sampling strategies for machine learning. For example, the training set of the model S1-1M-RS-30K contains 30K randomly sampled subscribers while that of the model S1-1M-P3 contains 36.7K subscribers with a 3% over-sampling of churners.

Table 4. Models Evaluation of Decision Tree without Segmentation

Model	Hit Ratio	Capture ratio	LIFT At 10%	Descriptlion
S1-1M-RS-30K	92.92%	85.91%	8.74	Analytical base=1M, Random Sample, Learning Records=30K
S1-1M-P3	97.90%	84.82%	9.18	Analytical base=1M, Over Sampling=3%, Learning Records=36.7K
S1-1M-P5	96.21%	94.55%	9.96	Analytical base=1M, Over Sampling=5%, Learning Records=22K
S1-1M-P10	96.72%	93.82%	9.93	Analytical base=1M, Over Sampling=3%, Learning Records=11K
S1-2M-RS	87.85%	92.95%	9.21	Analytical base=2M, Random Sample, Learning Records=30K
S1-3M-RS	98.04%	86.27%	9.53	Analytical base=3M, Random Sample, Learning Records=30K

The churn rate of whole population is only 0.71%. We over-sample the churners to get a higher churn rate training set for machine learning. Table 4 summarizes the results of six different sampling strategies for machine learning. For example, the training set of the model S1-1M-RS-30K contains 30K randomly sampled subscribers while that of the model S1-1M-P3 contains 36.7K subscribers with a 3% over-sampling of churners.

Table 4 shows that model S1-1M-P5 has the highest LIFT(10%), about 10. It means that the performance of using one-month analytical window to predict the possible churners of next month will be better than others.

2. Decision Tree Modeling With Customer Segmentation
Since all customers have been assigned to a cluster identity, we separate the training sets into 5 sub-sets for Decision Tree modeling by the cluster identity. Then we used the same validation sets to evaluate all models. Table 5 shows the performance of each clusters' decision tree model.

Table 5. Models Performance of Decision Tree with Segmentation

Predict Window	DT With Segment (C1)			DT With Segment (C2)			DT With Segment (C3)			DT With Segment (C4)			DT With Segment (C5)		
	Hit %	Cap%	LIFT	Hit %	Cap%	LIFT	Hit %	Cap%	LIFT	Hit %	Cap%	LIFT	Hit %	Cap%	LIFT
200108	79.91%	94.61%	9.75	64.86%	88.89%	9.89	78.09%	89.05%	9.47	46.93%	88.42%	8.95	33.07%	98.80%	19.02
200109	42.41%	93.99%	9.67	49.66%	92.44%	10.08	43.60%	90.23%	9.56	31.68%	94.32%	9.66	26.70%	98.32%	19.02
200110	32.23%	94.43%	9.69	36.86%	96.99%	9.55	33.98%	90.51%	9.64	15.76%	92.45%	9.43	12.90%	96.15%	18.87
200111	24.48%	95.32%	9.72	24.74%	92.00%	9.55	25.03%	91.09%	9.47	14.79%	93.10%	9.48	12.43%	95.65%	18.85
200112	21.67%	93.29%	9.53	21.85%	91.55%	9.87	19.15%	91.42%	9.60	9.88%	95.24%	9.76	4.32%	95.24%	18.82
200201	15.26%	93.39%	9.51	14.84%	90.68%	9.82	13.52%	2.31%	9.67	8.88%	84.75%	8.81	17.79%	89.47%	18.34
200202	2.34%	0.72%	1.06	0.00%	0.00%	0.94	9.90%	2.34%	1.27	11.58%	10.28%	1.08	4.30%	0.54%	10.96
200203	4.80%	1.39%	1.18	0.00%	0.00%	0.94	6.00%	1.35%	1.04	10.71%	8.11%	1.25	1.20%	2.42%	11.12
200204	5.88%	1.61%	1.29	2.56%	1.52%	1.08	12.77%	2.31%	1.20	13.33%	12.20%	2.06	0.30%	0.60%	10.97
200205	1.77%	0.42%	0.97	0.00%	0.00%	1.16	4.65%	0.60%	0.36	0.00%	0.00%	2.26	0.61%	1.16%	11.01
200206	1.79%	0.40%	1.33	0.00%	0.00%	0.94	14.89%	1.55%	0.64	4.41%	1.11%	2.16	0.00%	0.00%	10.92
200207	7.02%	1.63%	1.06	5.00%	1.82%	1.11	14.29%	1.40%	1.50	4.76%	1.11%	1.69	0.00%	0.00%	10.92

4.4.2. Neural Network (Back Propagation Network, BPN)
Based on other research results (e.g., Cybenco 19978; Zhan et al. 1998), we know that using one-layer hidden layer and optimal network design might get a more accurate model from Neural Network. In this study, we use 1-1-1 (input-hidden-output) as a training model type. This training type includes 43 inputs and only one output. Since public information is not available on key modeling parameters such as the learning rate or number of neurons in the

hidden layer, we try many different combinations. Table 6 shows the results, in which model N18-R6, for example, uses 18 neurons in the hidden layer with 0.6 learning rate.

To minimize other variances, we use the same training set for BPN as for Decision Tree. Table 6 shows that N21-R6 achieves the best performance from R-square and MSE measurements.

Table 6. Learning Results of BPN

	N18-R6	N19-R6	N20-R6	N21-R6	N22-R6	N23-R6	N24-R6	N25-R6
R squared	0.9934	0.9922	0.9998	0.9999	0.9998	0.9998	0.9995	0.9941
r squared	0.9934	0.9922	0.9998	0.9999	0.9998	0.9998	0.9995	0.9941
MSE	0.001	0.002	0	0	0	0	0	0.001

4.5. Model Performance Stability
4.5.1. Overall Performance Trend

We use the data from the telecom operator to "track" model performance over a period of time. Figure 5 shows the trend of performance in terms of Hit Rate and Capture Rate:

- Figure 5 shows that all the models demonstrate stable accuracy in the first 6 months. However, there a significant degradation occurs in the month of February 2002, regardless of modeling techniques. It is possible that the Chinese New Year is in February 2002 and consumers behave differently during Chinese New Year period.

- In the first 6 months, NN outperforms DT and DT without segmentation slightly outperforms DT with segmentation. In general, the performance of building a predictive model on individual segment is more accurate than the one built on the entire customer population. Thus, a decision tree model without segmentation should out-perform a decision tree model with segmentation. Our experiment shows otherwise. Furthermore, we are concerned about the significant performance gap after the first month between the NN and DT techniques.
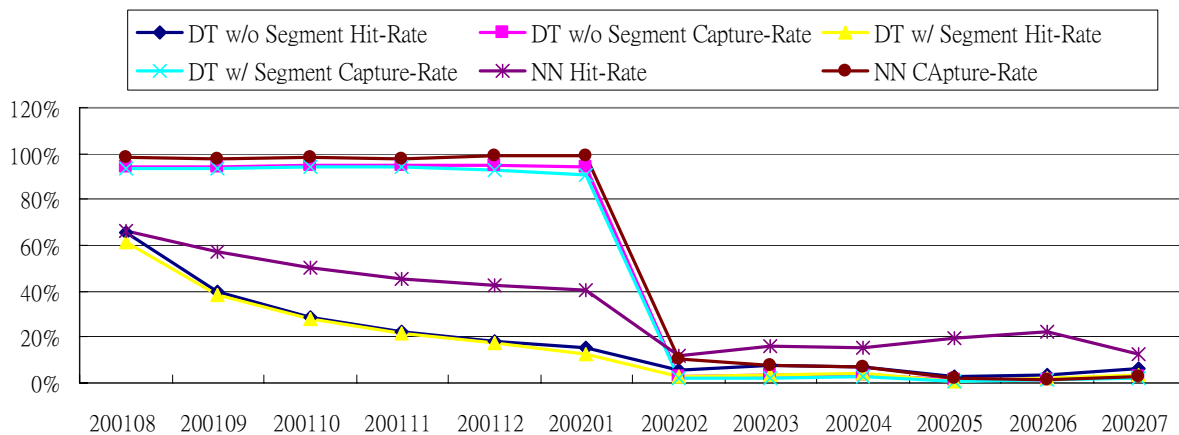

Figure 5: Hit Ratio & Capture Rate of techniques

4.5.2. Test Modeling Technique Differences

We use T-test to compare modeling techniques under different modeling parameters. The hypotheses are:

- H01: Hit Ratio of the decision tree model without segmentation isn't different from that

with segmentation. (DTH-DTSH)

- H02: Capture rate of the decision tree model without segmentation isn't different from that with segmentation. (DTC-DTSC)

- H03: Hit Ratio of the neural network model isn't different from the decision tree model without segmentation. (NNH-DTH)

- H04: Capture rate of the neural network model isn't different from the decision tree model without segmentation. (NNC-DTC)

- H05: Hit Ratio of the neural network model isn't different from the decision tree model with segmentation. (NNH-DTSH)

- H06: Capture rate of the neural network model isn't different from the decision tree model with segmentation. (NNC-DTSC)

Table 7 lists T-test results:
- The performance of decision tree model without segmentation is better than that with segmentation.

- The performance of BPN is better than decision tree model without segmentation on both hit ratio and capture rate.

Table 7. Model performance evaluation

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
DTH_DTSH	12	1.992E-02	1.146E-02	3.309E-03
DTC_DTSC	12	1.108E-02	8.062E-03	2.327E-03
NNH_DTH	12	.1508	8.284E-02	2.391E-02
NNC_DTC	12	3.333E-02	2.103E-02	6.072E-03
NNH_DTSH	12	.1708	7.416E-02	2.141E-02
NNC_DTSC	12	4.417E-02	2.610E-02	7.534E-03

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
DTH_DTSH	6.020	11	.000	1.992E-02	1.263E-02	2.720E-02
DTC_DTSC	4.762	11	.001	1.108E-02	5.961E-03	1.621E-02
NNH_DTH	6.307	11	.000	.1508	9.820E-02	.2035
NNC_DTC	5.490	11	.000	3.333E-02	1.997E-02	4.670E-02
NNH_DTSH	7.980	11	.000	.1708	.1237	.2180
NNC_DTSC	5.863	11	.000	4.417E-02	2.759E-02	6.075E-02

Note: DTH: Hit Raito of decision tree model without segmentation
DTC: Capture Rate of decision tree model without segmentation
DTSH: Hit Raito of decision tree model with segmentation
DTC: Capture Rate of decision tree model with segmentation
NNH: Hit Raito of Neural Network (BPN)

4.5.3. *Sample Size Impact*

One theory is that our results are biased because of limited churn samples in the analysis base: The mobile service provider only budgeted this study at the population of about 160,000 customers, and the associated monthly churn rate was only 0.71%. The data size was not sufficient to build a good predictive model by each customer segment because we could not explore real significant information from only few churners in each customer segment. For example, Table 3 shows that C3 contains about 17% of the customer population with a 0.3% churn rate. Thus, there are only 160,000 x 17% x 0.3%, or about 80 churners in C3.

4.5.4. *Robust Models*

In order to validate models' accuracy, we randomly selected 50,000 subscribers to evaluate models' performance. Figure 6 shows the results: Models built are more robust and performance disparity between NN and DT models in the first 6 months disappeared.
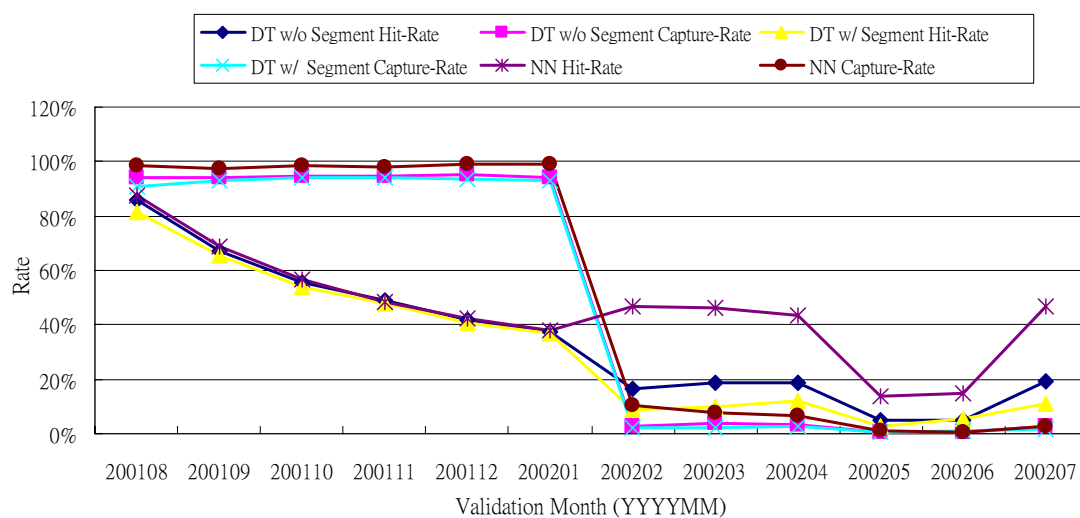

Figure 6: Model Performance with 50,000 Subscribers

Figure 7 compares the LIFT of all models: Both NN and DT techniques generate models with a hit rate of 98% from the top 10% of predicted churners in the list. That is, LIFT(10%) is about 10.
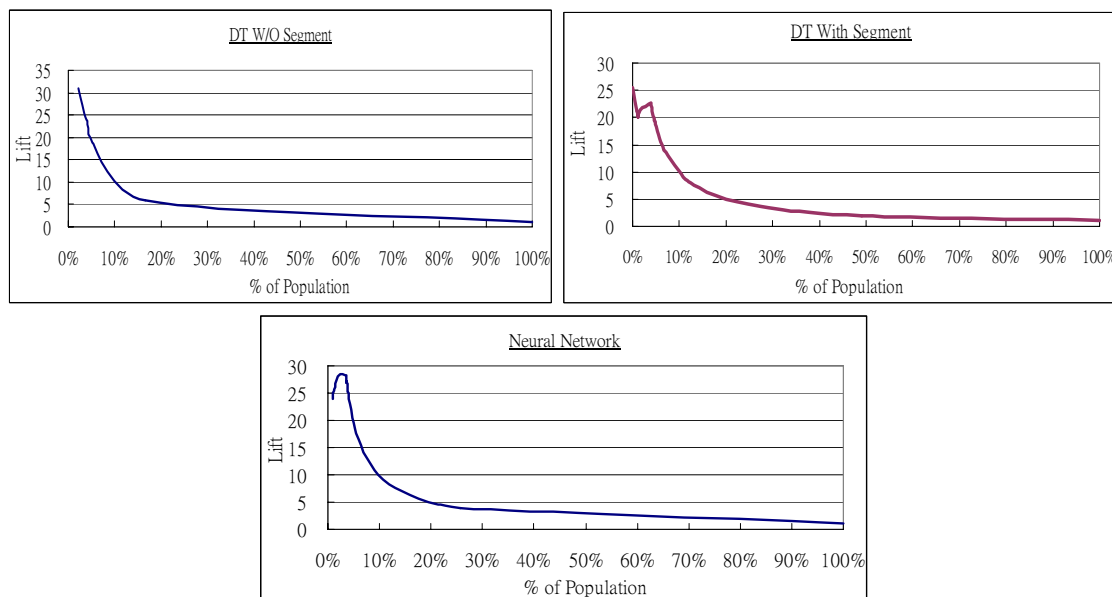

1135

Figure 7: LIFT Comparison

4.5.5. *Performance Comparison with Prior Studies*
Berson (Berson et al. 2000) used customer demographics, contractual data, and customers' service data as predictors to generate models with a Hit Rate of about 40% from the top 10% of predicted churners in the list. (That is, LIFT(10%) is about 4.) Wei (Wei et al. 2000) used customer call detail records as predictor and generated models with a Hit Rate of less than 50% from the top 10% of predicted churners in the list. (That is, LIFT(10%) is less than 5.) Our LIFT(10%) is about 10. Although the customer bases are different and there are other modeling parameters to consider, the LIFT achieved by all proposed techniques in this study demonstrates a significant improvement from those in early studies.

5. Conclusions and Future Research
Churn prediction and management is critical in liberalized mobile telecom markets. In order to be competitive in this market, mobile service providers have to be able to predict possible churners and take proactive actions to retain valuable customers.

In this research, we proposed different techniques to build predictive models for telecom churn prediction. We included customer service and customer complain log for modeling, as suggestions from prior researches of Wei et al. (Wei et al. 2000). We examine the impact of inadequate data on model building. Our empirical evaluation shows that data mining techniques can effectively assist telecom service provides to make more accurate churner prediction.

However, effective churn prediction model only supports companies to know which customers are about to leave. Successful churn management must also include effective retention actions. Mobile service providers need to build up attractive retention programs to satisfy those customers. Furthermore, integrating churn score with customer segment and applying customer value also helps mobile service providers to design the right strategies to retain high valuable customers.

Data mining techniques can be applied in many fields in the CRM space, such as credit card fraud detection, credit score, affinity between churners and retention programs, response modeling, and customer purchase decision modeling. We expect to see more data mining applications in business management, and more sophisticated data mining techniques will be developed as business complexity increases.

References
Berson, Alex, Smith, Stephen and Thearling, K. "Building Data Mining Applications for CRM", McGraw-Hill, New York, NY, 2000.
Cybenco, Horink "Approximation by Super-positions of Sigmoidal Function", Mathematical Control Cignal Systems, Vol 2, 1998, pp. 303-314.
Kentrias, Stelios "Customer Relationship Management-The SAS Perspective", www.cm2day.com, 2001.
Lejeune, Miguel A.P.M. "Measuring The Impact of Data Mining on Churn Management", Internet research: Electronic Network Applications and Policy Vol. 11, No. 5, 2001, pp. 375-387.
Mattersion, Rob. "Telecom Churn Management", APDG Publishing, NC, 2001.
SAS Institute "Best Price In Churn Prediction", a SAS institute white paper, 2000.

Setiono, Rudy, Liu, Huan "Neural-Network feature selector", IEEE transaction on neural network, Vol. 8(3), 1997, pp654-661.

Thearling, Kurt "A Introduction of Data Mining", Direct Marketing Magazine, Feb 1999.

Wei, Chih-Ping, Chiu, I-Tang "Tuning Telecommunications Call Detail to Churn Prediction: a Data Mining Approach', Expert System with Application 23, 2002, pp.103-112.

Zhan, G, Patuwo, B.E. and Hu, M.Y. "Forecasting with Artificial Neural Network: The state of the Art", International Journal of forecasting, Vol. 14, 1998, pp.35-62.