PERGAMON

Expert Systems with Applications 23 (2002) 103–112

Expert Systems with Applications

www.elsevier.com/locate/eswa

Turning telecommunications call details to churn prediction: a data mining approach

Chih-Ping Wei[a,*], I-Tang Chiu[b]

[a]*Department of Information Management, National Sun Yat-Sen University, Kaohsiung, Taiwan, ROC*
[b]*Southern Taiwan Business Group, Chunghwa Telecom Co. Ltd, Kaohsiung, Taiwan, ROC*

Abstract

As deregulation, new technologies, and new competitors open up the mobile telecommunications industry, churn prediction and management has become of great concern to mobile service providers. A mobile service provider wishing to retain its subscribers needs to be able to predict which of them may be at-risk of changing services and will make those subscribers the focus of customer retention efforts. In response to the limitations of existing churn-prediction systems and the unavailability of customer demographics in the mobile telecommunications provider investigated, we propose, design, and experimentally evaluate a churn-prediction technique that predicts churning from subscriber contractual information and call pattern changes extracted from call details. This proposed technique is capable of identifying potential churners at the contract level for a specific prediction time-period. In addition, the proposed technique incorporates the multi-classifier class-combiner approach to address the challenge of a highly skewed class distribution between churners and non-churners. The empirical evaluation results suggest that the proposed call-behavior-based churn-prediction technique exhibits satisfactory predictive effectiveness when more recent call details are employed for the churn prediction model construction. Furthermore, the proposed technique is able to demonstrate satisfactory or reasonable predictive power within the one-month interval between model construction and churn prediction. Using a previous demographics-based churn-prediction system as a reference, the lift factors attained by our proposed technique appear largely satisfactory. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Data mining; Telecommunications data mining; Churn prediction; Churn management; Classification analysis; Decision tree induction; Multi-classifier class-combiner approach

1. Introduction

As deregulation, new technologies, and new competitors have opened up the telecommunications industry, the telecommunications service market has become more competitive than ever (Gerpott, Rams, & Schindler, 2001; Kappert & Omta, 1997). To survive or maintain an advantage in an ever-increasing competitive marketplace, many companies are turning to data mining techniques to address such challenging issues as fraud detection (Burge & Shawe-Taylor, 1997; Cox, Eick, & Wills, 1997; Ezawa & Norton, 1996; Taniguchi, Haft, Hollmén, & Tresp, 1998), prospect profiling (Kappert & Omta, 1997; Wei, Chang, & Lee, 2000), churn prediction and management (Berson, Smith, & Thearling, 2000), etc.

Churn prediction and management is a concern for many industries, but it is particularly acute in the strongly competitive and now broadly liberalized mobile telecommunications industry. Subscriber churning (often referred to as customer attrition in other industries) in mobile telecommunications refers to the movement of subscribers from one provider to another. Many subscribers frequently churn from one provider to another in search of better rates/services or for the benefits of signing up with a new carrier (e.g. such as receiving the latest cellular phone). It is estimated that the average churn rate for the mobile telecommunications is 2.2% per month (Berson et al., 2000). That is, about 27% of a given carrier's subscribers are lost each year, making it essential to develop an effective churn-reduction method. The cost of acquisition of a new mobile service subscriber is estimated to be from $300 to $600 in sales support, marketing, advertising, and commissions (Berson et al., 2000; SPSS, 1999). However, the cost of retaining an existing subscriber is generally much lower than that. On the other hand, existing subscribers tend to generate more cash flow and profit, since they are less sensitive to price and often lead to sales referrals (Eiben, Euverman, Kowalczyk, & Slisser, 1998). Due to the high cost of acquiring new subscribers and considerable benefits of retaining existing ones, building a churn prediction model

* Corresponding author. Tel.: +886-7-5252-000 ext 4729; fax: +886-7-525-4799.
E-mail address: cwei@mis.nsysu.edu.tw (C.-P. Wei).

0957-4174/02/$ - see front matter
PII: S0957-4174(02)00030-1

to facilitate subsequent churn management and customer retention is critical for the success or bottom-line survival of a mobile telecommunications carrier in this greatly compressed market-space.

Data mining refers to a process of extracting previously unknown, valid and actionable patterns or knowledge from large databases for crucial business decision support (Berry & Linoff, 1997; Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998; Chen, Han, & Yu, 1996; Frawley, Piatetsky-Shapiro, & Matheus, 1991). Based on the kinds of knowledge which can be discovered in databases, data mining techniques can be broadly classified into several categories, including classification, clustering, dependency analysis, data visualization, and text mining (Shaw, Subramaniam, Tan, & Welge, 2001). Classification analysis is a process that induces a model to categorize a set of pre-classified instances (called training examples) into classes. Such a classification model is then used to classify future instances. Widely adopted classification techniques include decision tree induction, decision rule induction, and neural network. Clustering analysis is a process whereby a set of instances (without a predefined class attribute) is partitioned (or grouped) according to some distance metric into several clusters in which all instances in one cluster are similar to each other and different from the instances of other clusters. Dependency analysis discovers dependency patterns (e.g. association rules, sequential patterns, temporal patterns, and episode rules) embedded in data. Data visualization allows decision makers to view complex patterns in the data as visual objects in three dimensions and color; it supports advanced manipulation capabilities to slice, rotate or zoom the objects to provide varying levels of details of the patterns observed. Finally, text mining (including text categorization, document clustering, term association discovery, information extraction, etc.) extracts patterns from textual documents and can be applied to facilitate document management and retrieval or to discover knowledge hidden in texts.

Past research on churn prediction in the telecommunications industry mainly employed classification analysis techniques for the construction of churn prediction models, using as predictors (i.e. input variables) user demographics, contractual data, customer service logs and/or call patterns aggregated from call details (e.g. average call duration, number of outgoing calls, etc.). For example, the classification and regression trees (CART) algorithm (Breiman, Friedman, Olshen, & Stone, 1984) was employed for churn prediction based on customer demographics and contractual data (e.g. length of service, contract type, etc.) as well as customer service logs from the customer service center that captured inbound calls from the customers (Berson et al., 2000). However, existing churn-prediction systems have several disadvantages. First, use of customer demographics in churn prediction renders the resulting churn analysis at the customer rather than contract (or subscriber) level. In other words, propensities of churning are calculated on a per customer rather than contract basis. It is quite common that a customer concurrently holds several mobile service contracts with a particular carrier, with some contracts more likely to be churned than others. In this regard, customer-level-based churn prediction is considered inappropriate. Second, information on some of the input variables employed by existing churn-prediction systems frequently are not readily available. For example, the mobile telecommunications company investigated in this study has very limited customer demographic information (e.g. only the name, date of birth, identification number, and billing address were collected for each subscriber). Unavailability of customer profiles, prevailing in most telecommunications companies in many countries, limits the applicability of existing churn-prediction systems.

In response to the described limitations of existing churn-prediction systems, we exploit the use of call pattern changes and contractual data for developing a churn-prediction technique that identifies potential churners at the contract level. Conceivably, subscriber churn is not an instantaneous occurrence that leaves no trace. Before an existing subscriber churns, his/her call patterns might be changed (e.g. the number of outgoing calls gradually get reduced). In other words, changes in call patterns are likely to include warning signals pointing toward churning. Such call pattern changes can be extracted from subscribers' call details and are valuable for constructing a churn prediction model based on a classification analysis technique.

The remainder of the paper is organized as follows, Section 2 details the data and variables used for the target churn prediction problem. Section 3 depicts the proposed churn-prediction technique, using a decision tree induction algorithm for learning. Section 4 describes the evaluation design and discusses important experimental results. This paper is concluded in Section 5 with a summary, discussion of its contributions and limitations, and some future research directions.

2. Data and variables

The mobile telecommunications company investigated in this study is one of the largest providers in Taiwan. As of November 2001, the carrier had a nationwide customer base of some 21,000,000 mobile service subscribers. Although its average churn rate is below the industry average in Taiwan, the company still is experiencing a high churn rate of 1.5–2% monthly.

As mentioned, the investigated company collects and maintains very limited information about its subscribers. An analysis of its operational databases suggests that two types of readily available data for target churn prediction include contractual data and the call details of subscribers. Specifically, for each subscriber (existing or prior), the contractual information consists of its phone number, contract type, payment type, contract starting date, and contract termination date (if the subscriber has ended this contract). Based on interviews with senior sales representatives of the investigated company, contract-related variables potentially affecting a subscriber's propensity to churn were identified as follows:

1. Length of services (LOS): For a previous subscriber, LOS refers to the duration, measured in days, between the contract starting date and termination date; otherwise, it measures the number of days since the contract started. Conceivably, if a subscriber's relationship with a company is longer, he/she will less likely become a churner.
2. Payment type: This refers to whether or not a subscriber authorizes direct payment of mobile service bills from his/her bank account. Perhaps, a subscriber with the direct-payment option is more likely to stay with the company longer.
3. Contract type: Different contract types impose different subscription terms and rates. In this study, we only differentiate between whether or not a contract requires a deposit, which often creates a disincentive to switching between carriers.

On the other hand, the call details maintained in the investigated company for billing purposes include for each outgoing call such fields as the caller number, the receiver's number, the date, starting time, ending time and duration of the call, as well as the charge applied to this call. Three measures commonly used to describe the call patterns of a subscriber by aggregating his/her recall records consist of:

1. Minutes of use (MOU): referring to the total number of minutes of the outgoing calls made by the subscriber over a specific period.
2. Frequency of use (FOU): referring to the total number of outgoing calls made by the subscriber over a specific period.
3. Sphere of influence (SOI): referring to the total number of distinct receivers contacted by the subscriber over a specific period.

To represent call pattern changes of a subscriber during a specific observation period (T), T is divided into several consecutive sub-periods of equal duration. Assuming T is divided into n sub-periods, we modeled the call pattern changes of a subscriber by considering the change rate of each measure between any two consecutive sub-periods. Specifically, in this study, the variables used for representing call pattern changes of a subscriber included:

1. $\text{MOU}_{\text{initial}}$: denotes the MOU of a subscriber in the first sub-period.
2. $\text{FOU}_{\text{initial}}$: denotes the FOU of a subscriber in the first sub-period.
3. $\text{SOI}_{\text{initial}}$: denotes the SOI of a subscriber in the first sub-period.
4. ΔMOU_s: denotes the change in MOU of a subscriber between the sub-period $s-1$ and s (for $s = 2, \ldots, n$) and is measured by $\Delta\text{MOU}_s = (\text{MOU}_s - \text{MOU}_{s-1} + \delta)/(\text{MOU}_{s-1} + \delta)$, where $\text{MOU}_1 = \text{MOU}_{\text{initial}}$ and δ is a small positive real number (e.g. 0.01) to avoid the case when MOU_{s-1} is 0 (i.e. when ΔMOU_s cannot be calculated).
5. ΔFOU_s: denotes the change in FOU of a subscribe between the sub-period $s-1$ and s (for $s = 2, \ldots, n$) and is calculated as $\Delta\text{FOU}_s = (\text{FOU}_s - \text{FOU}_{s-1} + \delta)/(\text{FOU}_{s-1} + \delta)$.
6. ΔSOI_s: denotes the change in SOI of a subscriber between the sub-period $s-1$ and s (for $s = 2, \ldots n$) and is calculated as $\Delta\text{SOI}_s = (\text{SOI}_s - \text{SOI}_{s-1} + \delta)/(\text{SOI}_{s-1} + \delta)$.

The number of variables used for constructing a churn-prediction model varies with the number of sub-periods used for representing call pattern changes over a specific observation period. Imaginably, a classification algorithm using a greater number of sub-periods might capture subtle changes in call patterns potentially leading toward churning; thus, possibly improving its churn prediction accuracy. However, as the number of sub-periods increases, the duration of each sub-period will decrease. As a result, a classification algorithm using a greater number of sub-periods might be more susceptible to variations in call patterns, possibly impairing its generalizability and predictive effectiveness. We will examine experimentally the effects of the number of sub-periods on the churn prediction effectiveness in Section 4.

3. Development of churn prediction technique

Given a set of subscribers (known as training instances), each of which is described by the input variables depicted previously and labeled to indicate the user's churn status (generally referred to as a decision outcome in the classification analysis technique), a classification-analysis-based churn-prediction technique induces a churn prediction model (i.e. establishes relationships between churn status and variables) from the set of training instances. Depending on the learning strategies adopted as well as the types of classification models induced, several classification analysis approaches have emerged and received fairly extensive research attention, including decision tree (e.g. ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993)), decision rule (e.g. AQ family (Michalski, Mozetic, Hong, & Lavrac, 1986) and CN2 (Clark & Boswell, 1991; Clark & Niblett, 1989)), and neural network (e.g. back propagation neural networks (Rumelhart, Hinton, & Williams, 1986)). The decision tree approach induces a tree-based classification model to describe relationships between variables and decision outcomes and uses the resulting decision tree for subsequent prediction purposes. On the other hand, the decision rule approach examines relationships between variables and outcome decisions to discover a set of decision rules (e.g. ordered or unordered) rather than a tree-based classification model. Meanwhile, the neural network approach uses a predetermined network topology to produce a set of adequately weighted links, which jointly differentiate individual decision outcomes, based on the respective variable values.

To a great extent, a neural network technique represents a holistic approach to learning by encoding the classification model in the weights between nodes. Thus, its resulting knowledge often lacks interpretability, diminishing the attractiveness of its application to churn prediction and management. Furthermore, a neural network requires a long training time due to its iterative nature, making it impractical for churn prediction in the mobile telecommunications industry where the subscription base is often large. Consequently, the decision tree and the decision rule approaches appear to be more appropriate for targeted learning and prediction, because they are capable of efficiently generating interpretable knowledge in an understandable form; i.e. a decision tree or decision rules. In this study, we took the decision tree approach (specifically C4.5 (Quinlan, 1993)) as the basis for the development of our churn-prediction technique. Choice of the decision-tree over the decision-rule approach primarily was based on its popularity and comparable learning effectiveness.

As mentioned, the described churn prediction application exhibits a highly skewed data in decision outcomes. Specifically, as mentioned previously, a randomly selected dataset from the company investigated in this study included 1.5–2% churners and 98–98.5% non-churners. Such a highly skewed distribution problem, if not properly addressed, would imperil the resulting learning effectiveness and might result in a 'null' prediction system that simply predicts all instances as having the majority decision class as the training instances (e.g. predicting all subscribers as non-churners).

In this study, we adopted the multi-classifier class-combiner approach proposed by Chan, Fan, Prodromidis, and Stolfo (1999) to address the highly skewed class distribution problem. This approach simulates real-world practices where multiple experts, typically imperfect, are engaged in a decision making process for making individual recommendations which jointly are then consolidated by another expert or individual into a final decision. For a dichotomous classification problem, the multi-classifier class-combiner approach proceeds in the following manner. Given a set of training instances S, the multi-classifier class-combiner first creates from S multiple training subsets with a desired class ratio, where the instances having the majority decision class (referred to as the majority instances) are evenly and randomly partitioned into training subsets and the instances belonging to the minority decision class (referred to as the minority instances) are replicated across these training subsets. For example, assume the ratio between the minority and the majority instances in a training set to be 1:10 and the desired ratio in a training subset to be 1:1. The majority instances in the original training set are randomly and evenly divided into 10 training subsets, each of which will include all of the minority instances. Thus, the ratio between the minority and the majority instances in each resulting training subset becomes 1:1. Formally, let N be the size of a training set having the ratio between the minority and majority instances in the training set being $1{:}x$, and $1{:}y$ being the desired ratio in each training subset. Hence, the number of training subsets is x/y, where each training subset has $N/(1+x)$ minority instances and $(N \times y)/(1+x)$ majority instances.

Subsequently, for each training subset a classification model (or base classifier) is generated by a classification analysis technique. To predict the decision outcome of an unseen instance, each base classifier individually processes this instance. An overall prediction for the unseen instance is then delivered by a meta-classifier which combines the predictions made by the individual base classifiers. In this study, we employed a weighted voting-based strategy for the meta-classifier. Assume the dichotomous classification problem consisting of decision classes C_1 and C_2. Let C_1 be assigned a weight w_1 and C_2 be assigned another weight w_2 (i.e. $1 - w_1$). Let n_{c_1} and n_{c_2} be the number of base classifiers predicting C_1 and C_2, respectively, as the decision outcome for the unseen instance. Thus, the meta-classifier predicts the decision C_1 for the unseen instance if $w_1 \times n_{c_1} > w_2 \times n_{c_2}$; otherwise, the decision C_2.

4. Empirical evaluation

Call details collected from a Southern Taiwan branch of the mobile telecommunications company were investigated. Specifically, the call records made between October 2000 and January 2001 were included in this evaluation. The contractual data of all subscribers were also obtained. We excluded from the evaluation those subscribers whose mobile services were terminated by the company during this data collection period because their payments were delinquent. As a result, the customer base consisted of some 114,000 subscribers; 4500 of them disconnected their mobile services with the company across this data collection period. The total number of call records available for the evaluation was about 9,100,000.

4.1. Evaluation design and procedure

Using the collected source data set, we randomly selected a prediction period (P) to generate an evaluation data set as well as to determine the churn status (i.e. the decision outcome) for each subscriber included in this evaluation data set. The churn status of a subscriber was defined as the connected or disconnected status of the subscriber within the prediction period P. A subscriber disconnecting his/her mobile service with the investigated company during P was considered a churner. If, at the end of P, a subscriber still had service with the investigated company, he/she would be classified as a non-churner. Any subscriber discontinuing cellular phone service before P was not included in this evaluation data set.

Moreover, churn prediction is not the last task of churn management. Once a set of potentially at-risk subscribers is identified, a mobile telecommunications provider needs some time to approach and focus its retention actions on these subscribers before they become churners. In this light, a retention period (R) was specified immediately prior to the prediction period for such customer retention endeavors and, thus, the call records during the retention period would not be used for churn-prediction model construction purposes. Prior to R, we designated an observation period (T) from which the call records were employed for extracting the call pattern changes of subscribers to predict churning. Any subscriber whose contract started no earlier than the observation period T was excluded from this evaluation data set. In essence, given a data set prepared in the described manner, the evaluation design was intended to examine the predictive effectiveness of the proposed technique employing call details (together with contractual information) of subscriber usage in the observation period T and for predicting their status in the prediction period P; thereby practically giving the investigated company a retention period R for focusing their retention actions on the potential churners predicted by the proposed technique.

By its inductive nature, the proposed churn-prediction technique requires learning and testing. Thus, an evaluation data set needs to be divided exclusively into training or testing instances. The proposed technique induces from the set of training instances a classification model that is then tested using another set (i.e. testing instances). A multiple learning-and-testing evaluation approach is preferred because a single learning-and-testing experiment may generate misleading predictive effectiveness estimates. Consequently, we chose a 10-fold cross validation approach (Weiss & Kulikowski, 1991). With this particular approach, the instances in an evaluation data set were randomly divided into ten mutually exclusive data subsets of approximately equal size. In each learning-and-testing process, one data subset was chosen as the testing set and the others were used for learning. The learning-and-testing process was performed ten times and a performance estimate was calculated by averaging the testing performance that was recorded in each separate learning-and-testing session.

Furthermore, to avoid potential effects of internal and external event occurrences (e.g. promotion activities by other carriers) during the observation period and the retention period of an evaluation data set on the evaluation validity, we generated three evaluation data sets by randomly selecting three different prediction periods. Each of the data sets was evaluated using the 10-fold cross validation approach. An overall performance estimate over the three data sets was calculated by averaging the testing performance that was recorded for each data set.

4.2. Evaluation criteria

In this study, the predictive effectiveness of a churn-prediction technique is measured by the miss and false alarm rates. The miss rate is defined as the percentage of that a churn-prediction technique fails to detect a churner, while the false alarm rate is defined as the percentage of the non-churners that are incorrectly predicted as churners by a churn-prediction technique. A detection error tradeoff (DET) curve (Yang, Pierce, & Carbonell, 1998) is employed to address the inevitable tradeoffs between miss and false alarm rates. A churn-prediction technique with its DET curve closer to the origin (i.e. perfect prediction) would be more desirable. A high false alarm rate increases the unnecessary cost for retaining loyal subscribers, while a high miss rate increases the opportunity cost of not retaining at-risk subscribers. Because the cost of retaining a subscriber is much lower than that of acquiring a new subscriber in the telecommunications industry, reducing the opportunity cost is clearly more important than lowering the unnecessary retention cost. In this light, a churn-prediction technique should aim at achieving the lowest attainable miss rate while maintaining false alarm rate at an acceptable level.

4.3. Effects of desired class ratios

As mentioned, the proposed churn-prediction technique adopted the multi-classifier class-combiner approach to address the problem of a highly skewed class distribution between churners and non-churners. The desired ratio between the minority instances (i.e. churners) and the majority instances (i.e. non-churners) determines the number of base classifiers and their corresponding training subsets used for churn prediction. We investigated the effects of desired class ratios on predictive effectiveness of the proposed churn-prediction technique. Specifically, the desired ratios of 1:2, 1:4, 1:8, and 1:16 between churners and non-churners were examined. The weight (as required by the weighted voting-based strategy in the meta-classifier) assigned to the non-churner class, ranging from 0.01 to 0.99 at 0.01 increments (or equivalently, the weight assigned to the churner class, ranging from 0.99 down to 0.01 at 0.01 decrements) were examined. Also included in this evaluation was a single-classifier approach (constructed from an original, highly skewed training data set) which provided a performance benchmark.

In this evaluation, the prediction period (P) was set to 7 days, the retention period (R) was 14 days, and the observation period (T) was 30 days. The observation period in an evaluation data set was segmented into 2, 3, 5 or 10 sub-periods for modeling call pattern changes. We evaluated the effects of desired class ratios on predictive effectiveness across different numbers of subperiods and weighting-schemes. The resulting DET curves over the range of numbers of sub-periods were largely similar. Thus, only the DET curve that resulted from two sub-periods will be

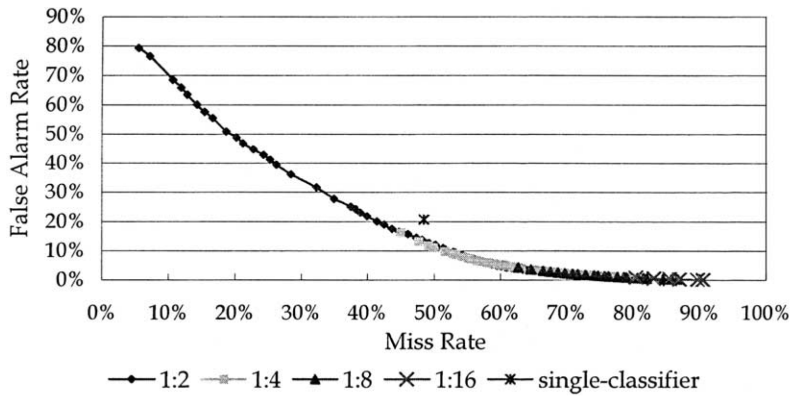

Fig. 1. Experimental results—effects of desired class ratios (using the number of sub-periods as 2).

shown and discussed later. As shown in Fig. 1, the DET curves attained by the multi-classifier class-combiner approach using different desired class ratios were largely overlapped. Evidently, the increase in the weight assigned to the non-churner class (i.e. toward favoring non-churner decisions) resulted in a decrease in the false alarm rate at the cost of miss rate. When compared with the single-classifier approach that resulted in a false alarm rate of 20.82% and a miss rate of 48.44%, the multi-classifier class-combiner approach across various desired class ratios appeared to demonstrate better tradeoffs between false alarm and miss rates; its resulting DET curves were closer to the origin (i.e. perfect prediction).

At the desired class ratio grew from 1:2 to 1:16, the proportion of non-churner instances increased in each training subset, resulting in over-optimistic base classifier behavior (i.e. increasing the tendency of predicting the subscribers in a test data set to be non-churners). As a result, the resulting DET curve shifted toward the area of low false alarm rate with high miss rate.

As the extreme when the desired class ratio was 1:16, the achieved false alarm rates across different weighting-schemes were lower than 1% while their respective miss rates were higher than 80%. However, when the desired class ratio of 1:2 was employed in the multi-classifier class-combiner approach, a prolonged DET curve was obtained. A longer DET curve would be more preferable since it would provide a larger range when trading-off between false alarm and miss rates. Hence, we selected the desired class ratio of 1:2 for use by the multi-classifier class-combiner approach in the subsequent evaluations.

4.4. *Effects of number of sub-periods*

We examined the effects of number of sub-periods (n) on the predictive effectiveness of the proposed technique while setting the desired class ratio (between churners and non-churners in each training subset) as 1:2, the prediction period as 7 days, the retention period as 14 days, and the observation period as 30 days. Given a selected observation period, various number of subperiods (specifically, $n = 2, 3, 5$, and 10) used for modeling call pattern changes of subscribers were investigated. As shown in Fig. 2, an increase in the number of sub-periods appeared to have shown no, or at most marginal, effects on churn prediction effectiveness since the resulting DET curves were highly comparable in length and trend. In other words, when a given observation period was segmented into more sub-periods (i.e. n increased), the capability to capture more subtle call pattern changes leading toward churning might be offset by increasing vulnerability to variations in call patterns that were irrelevant to churn predication. Fewer

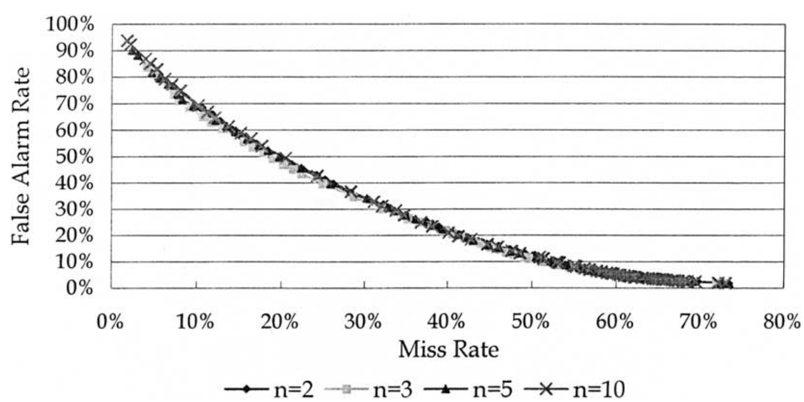

Fig. 2. Experimental results—effects of numbers of sub-periods.

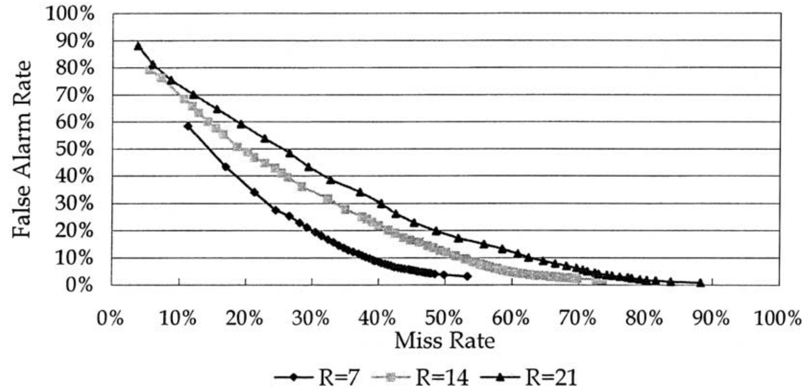

Fig. 3. Experimental results—effects of lengths of retention period.

sub-periods mean fewer input variables processed by the proposed churn-prediction technique. Accordingly, we selected 2 as the number of sub-periods for subsequent experiments.

4.5. Effects of lengths of retention period

As mentioned, a retention period (R), reserved between the observation and the prediction period, is designed to allow a mobile telecommunications company to take retention actions on those subscribers who are likely to churn during the predication period. An appropriate length of R is dictated by how efficiently a mobile telecommunications provider can formulate and execute a customer retention plan. In this experiment, we examined the effects of length of retention period, ranging from 7 to 21 at increments of 7, on the predictive effectiveness of the proposed churn-prediction technique. We set the prediction period as 7 days, the observation period as 30 days, the desired class ratio as 1:2, and the number of sub-periods as 2. As shown in Fig. 3, when the length of retention period was expanded from 7 to 21 days, its respective DET curve moved away from the perfect prediction. At any level of miss rate (or false alarm rate), a shorter retention period resulted in a lower false alarm (or miss rate). That is, more recent call behavior would be more useful for churn prediction. Such experimental results suggest that the interval between a subscriber's churning intention to his/her action would not be very long; thus, requiring a mobile telecommunications carrier to have responsive customer retention capability.

4.6. Analysis of temporal sensitivity

Imaginably, a subscriber's call behavior may evolve over time or be susceptible to events in the mobile telecommunications service market. A churn prediction model induced in the past might not be effective in the future. Thus, the temporal sensitivity of a churn prediction model constructed by the proposed technique needs to be evaluated. That is, using a churn prediction model built upon call records from a particular observation period, we intended to examine the predictive effectiveness of the model when it was applied to test data sets generated for different prediction periods. Since the source data obtained in this study consisted of only call records of 4 months, we could only investigate the predictive effectiveness of a given churn prediction model when applied to a test data set of the same period as, 1 or 2 months later than the training data set. As with the previous experiments, we set the prediction period as 7 days, the retention period as 14 days, the observation period as 30 days, the desired class ratio as 1:2, and the number of sub-periods as 2.

Specifically, three training data sets were randomly generated using respective prediction periods in between November 15 and 30, 2000. The three churn prediction models were then tested using the test data sets obtained from the same period, 1 or 2 months later. Fig. 4 shows the average prediction effectiveness, measured in false alarm and miss rates, over time. As shown, the predictive effectiveness of a churn prediction model appeared to decay gradually overtime. The proposed technique was able to exhibit satisfactory or reasonable predictive power when the interval between the model construction and churn prediction was within 1 month. The decay rate appeared to increase moderately as the interval further expanded. Based on these experimental results, it is suggested that, due to the evolving nature of subscribers' call behavior that, at the same time, is likely to be susceptible to events in the mobile telecommunications service market, churn prediction based on call pattern changes would require constant re-learning or re-discovery of an up-to-date churn prediction model. Thus, a corporate data warehouse that provides a subscriber-centric data repository by integrating multiple operational databases is essential to provide fast and easy data access for model construction and churn prediction purposes. Moreover, an efficient classification analysis algorithm (e.g. a decision tree induction algorithm adopted in this study) would be desirable to address such knowledge maintenance requirements.

4.7. Comparisons to previous studies

The churn prediction technique proposed in this study

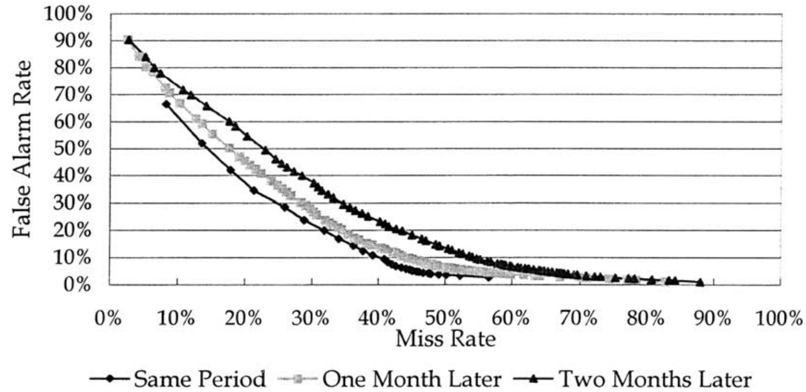

Fig. 4. Experimental results—temporal sensitivity of churn prediction model.

uses as predictors the contractual information of subscribers and their call pattern changes extracted from the call details. The proposed technique significantly differs from existing churn-prediction systems in the following three perspectives:

1. As mentioned, use of customer demographics in existing churn-prediction systems results in churn analysis at the customer level. On the contrary, the input variables employed by the proposed technique pertain to contracts rather than customer and is capable of identifying potential churners at the contract level.
2. The proposed technique does not rely on any customer demographics. Rather, it constructs a churn prediction model from contractual information and call pattern changes extracted from call details, thereby leveraging data already collected for other purposes. In this light, since the proposed technique can readily be applied to providers that have no or limited customer profiles, its applicability is higher than that of existing churn-prediction systems.
3. As mentioned, the proposed technique employs data observed in a period established for predicting subscribers' churn status in a prediction period, thus giving a mobile telecommunications provider a period to utilize its customer retention endeavor. Existing churn-prediction systems largely overlook such timing consideration, which contributes the essential linkage between its churn prediction and subsequent churn management and customer retention tasks.

The lift factors achieved by the proposed technique are shown in Fig. 5. The diagonal line represents an untargeted effort that randomly selects a subset of subscribers as potential churners. That is, an untargeted effort selects $s\%$ of the subscribers that will contain $s\%$ of the true churners; thus, resulting in a lift factor of 1. The proposed technique's predictive effectiveness depicted previously, can be transformed into a respective lift curve. Particularly, using a retention period of 7 days, the number of sub-periods of 2, and a desired class ratio of 1:2 (referred to as Model 1 in Fig. 5), the proposed technique was capable of identifying 10.03% of the subscribers that contained 54.33% of the true churners (i.e. a lift factor of 5.42), 20% of the subscribers that contained 64.72% of the true churners (i.e. a lift factor of 3.24), and 29.68% of the subscribers that contained 72.16% of the true churners (i.e. a lift factor of 2.43). On the other hand, using a retention period of 14 days, the number of sub-periods of 2, and a desired class ratio of 1:2 (referred to as Model 2 in

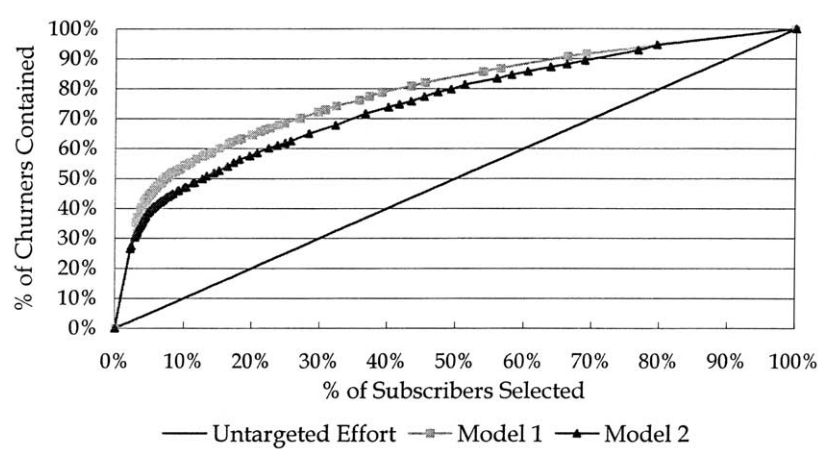

Fig. 5. Lift chart attained by the proposed churn-prediction technique.

Fig. 5), the proposed technique was able to identify 10.03% of the subscribers that contained 46.95% of the true churners (i.e. a lift factor of 4.68), 19.65% of the subscribers that contained 57.58% of the true churners (i.e. a lift factor of 2.93), and 28.32% of the subscribers that contained 65.07% of the true churners (i.e. a lift factor of 2.30). Evidently, in either model, the proposed churn-prediction technique appeared to achieve higher lift factors than the untargeted effort.

In contrast, a prior study using customer demographics, contractual data, and customer service data as predictors was able to identify 10.5% of the subscribers containing 41.5% of the true churners (i.e. a lift factor of 3.95) and 19.7% of the subscribers containing 55.8% of the true churners (i.e. a lift factor of 2.83) (Berson et al., 2000). Although a fair comparison with this prior study is difficult since the data sets are different, the lift factors attained by our proposed technique were considered largely satisfactory in view of those achieved by the prior study as a reference.

5. Conclusions and future research directions

Churn prediction and management is critical in the fast changing, strongly competitive and now broadly liberalized mobile communications market. To be able to improve customer retention, a mobile telecommunications service provider has to be able to predict at-risk subscribers on whom the subsequent customer retention effort is focused. In response to the limitations of existing churn-prediction systems and the unavailability of customer demographics in the mobile telecommunications provider studied, we proposed, designed, and experimentally evaluated a churn-prediction technique using as predictors the contractual information of subscribers and their call pattern changes extracted from the call details. The proposed technique is capable of identifying potential churners at the contract level for a specific prediction period. In addition, the proposed technique adopted the multi-classifier class-combiner approach to address the challenge of highly skewed class distribution between churners and non-churners. The empirical evaluation results suggested that multi-classifier class-combiner approach outperformed the single-classifier approach. The proposed call-behavior-based churn-prediction technique exhibited satisfactory predictive effectiveness when more recent call details were employed for the churn prediction model construction. Furthermore, the proposed technique was able to demonstrate satisfactory or reasonable predictive power within a one-month interval between model construction and churn prediction. Using a prior demographics-based churn-prediction system as a reference, the lift factors attained by our proposed technique were considered largely satisfactory.

This study benefits not only churn prediction research and practice but also other data mining applications with identical or similar characteristics. Continuing research should be aligned toward furthering the effectiveness and generalizability of the proposed technique. Some ongoing and future research directions are briefly summarized as follows. First, this study employed only the contractual data and call details for target churn prediction. The inclusion of additional input variables (e.g. extracting from customer complains and service logs) into our proposed technique might further enhance its predictive effectiveness. Second, subscribers in different geographical locations may exhibit dissimilar call behaviors. Thus, performing empirical evaluations of the developed technique in different geographical locations represents an interesting direction for further research. Third, as mentioned, constant re-learning or re-discovery of a churn prediction model is required due to the evolving nature of subscribes' call behavior that, at the same time, is likely to be susceptible to events in the mobile telecommunications service market. The provision of a subscriber-centric data warehouse would be desirable supporting the described knowledge maintenance requirement. In addition, churning is not restricted to the telecommunications market and is also a great concern for those industries (e.g. credit card issuers and internet service providers) where stiff competition provides incentives for customers to switch. Thus, expanding the developed technique to other industries suggests interesting directions for future research.

Acknowledgements

The authors wish to thank Yen-Hsien Lee for his technical support in data preparation.

References

Berry, M. J. A., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support*, New York: Wiley.

Berson, A., Smith, S., & Thearling, K. (2000). Customer retention. In *Building data mining applications for CRM*, New York: McGraw-Hill, Chapter 12.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*, Pacific Grove: Wadsworth.

Burge, P., & Shawe-Taylor, J. (1997). Detecting cellular fraud using adaptive prototypes. *Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management* (pp. 9–13). RI: Providence.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: From concept to implementation*, Upper Saddle River, NJ: Prentice-Hall PTR.

Chan, P. K., Fan, W., Prodromidis, A. L., & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems*, *14* (6), 67–74.

Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, *8* (6), 866–883.

Clark, P., & Boswell, R. (1991). Rule induction with CN2: Some recent improvements. *Proceedings of the Fifth European Conference (EWSL'91)* (pp. 151–163).

Clark, P., & Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning*, *3* (4), 261–283.

Cox, K. C., Eick, S. G., & Wills, G. J. (1997). Visual data mining:

Recognizing telephone calling fraud. *Journal of Data Mining and Knowledge Discovery*, *1* (2), 225–231.

Eiben, A. E., Euverman, T. J., Kowalczyk, E., & Slisser, F. (1998). Modelling customer retention with statistical techniques, rough data models, and genetics programming. In A. Skowron & S. K. Pal, *Fuzzy sets, rough sets and decision making processes,* Berlin: Springer.

Ezawa, K. J., & Norton, S. W. (1996). Constructing Bayesian networks to predict uncollectible telecommunications accounts. *IEEE Expert*, *11* (5), 45–51.

Frawley, W., Piatetsky-Shapiro, G., & Matheus, C. J. (1991). Knowledge discovery in databases: An overview. In G. Piatesky-Shapiro & W. J. Frawley, *Knowledge discovery in databases* (pp. 1–30). Cambridge, MA: AAAI/MIT Press, Chapter 1.

Gerpott, T. J., Rams, W., & Schindler, A. (2001). Customer retenion, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy*, *25* (10–11), 885–906.

Kappert, C. B., & Omta, S. W. F. (1997). Neural networks and business modeling—an application of neural modeling techniques to prospect profiling in the telecommunications industry. *Proceedings of the 30th Hawaii International Conference on System Science* (Vol. 5) (pp. 465–473).

Michalski, R. S., Mozetic, I., Hong, J., & Lavrac, N. (1986). The multipurpose incremental learning system AQ15 and its testing application to three medical domains. *Proceedings of American Association for Artificial Intelligence* (*AAAI-86*) (Vol. 2) (pp. 1041–1045).

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 106.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*, San Mateo, CA: Morgen Kaufmann.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland, *Parallel distributed processing: Explorations in the microstructures of cognition* (pp. 318–362). Vol. 1. MA: MIT Press.

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, *31* (1), 127–137.

SPSS Inc. (1999). Working with telecommunications: Churning in the telecommunications industry. *SPSS White Paper* (available at: http://www.spss.com/downloads/papers/).

Taniguchi, M., Haft, M., Hollmén, J., & Tresp, V. (1998). Fraud detection in communications networks using neural and probabilistic methods. *Proceedings of International Conference in Acoustics, Speech and Signal Processing* (*ICASSP'98*) (pp. 1241–1244).

Wei, C., Chang, H. T., & Lee, Y. H. (2000). Telecommunications data mining for target marketing. *Journal of Computers*, *12* (4), 60–74.

Weiss, S. M., & Kulikowski, C. A. (1991). *Computer systems that learn: Classification and prediction methods from statistics, neural nets, machine learning, and expert systems*, Los Atlos, CA: Morgan Kaufmann.

Yang, Y., Pierce, T., & Carbonell, J. (1998). A study on retrospective and on-line event detection. *Proceedings of 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 28–36) New York: ACM Press.