Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry

Michael C. Mozer, Richard Wolniewicz, David B. Grimes, *Student Member, IEEE*, Eric Johnson, and Howard Kaushansky

***Abstract*—Competition in the wireless telecommunications industry is fierce. To maintain profitability, wireless carriers must control *churn*, which is the loss of subscribers who switch from one carrier to another. We explore techniques from statistical machine learning to predict churn and, based on these predictions, to determine what incentives should be offered to subscribers to improve retention and maximize profitability to the carrier. The techniques include logit regression, decision trees, neural networks, and boosting. Our experiments are based on a database of nearly 47 000 U.S. domestic subscribers and includes information about their usage, billing, credit, application, and complaint history. Our experiments show that under a wide variety of assumptions concerning the cost of intervention and the retention rate resulting from intervention, using predictive techniques to identify potential churners and offering incentives can yield significant savings to a carrier. We also show the importance of a data representation crafted by domain experts. Finally, we report on a real-world test of the techniques that validate our simulation experiments.**

***Index Terms*—Boosting, churn, customer satisfaction, decision trees, logistic regression, prediction, profit maximization, retention, telecommunications, wireless industry.**

I. INTRODUCTION

COMPETITION in the wireless telecommunications industry is fierce. As many as seven competing carriers operate in each market. The industry is extremely dynamic, with new services, technologies, and carriers constantly altering the landscape. Carriers announce new rates and incentives weekly, hoping to entice new subscribers and to lure subscribers away from the competition. The extent of these rivalries is reflected in the deluge of advertisements for wireless service in daily newspapers and other mass media.

The United States had 69 million wireless subscribers in 1998: roughly 25% of the population. Some markets are further developed; for example, the subscription rate in Finland is 53%. Industry forecasts are for a U.S. penetration rate of 48% by 2003. Although there is significant room for growth in most markets, the industry growth rate is declining, and competition is rising. Consequently, it has become crucial for wireless carriers to control *churn*—the loss of customers who switch from one carrier to another. At present, domestic monthly churn rates are 2–3% of the customer base. At an average cost of $400 to acquire a subscriber, churn cost the industry nearly $6.3 billion in 1998; the total annual loss rose to nearly $9.6 billion when lost monthly revenue from subscriber cancellations is considered [5]. It costs roughly five times as much to sign on a new subscriber as to retain an existing one. Consequently, for a carrier with 1.5 million subscribers, reducing the monthly churn rate from 2% to 1% would yield an increase in annual earnings of at least $54 million and an increase in shareholder value of approximately $150 million. (Estimates are even higher when lost monthly revenue is considered; see [2] and [5].)

Typically, carriers attempt to control churn with a welcome call after subscribers have been on board for one or two months. Churn rates among subscribers at this point are about 10%. Thus, carriers have concluded that it is profitable to operate a call center that contacts nine satisfied subscribers for every one dissatisfied subscriber. Until very recently, there was seldom, if ever, a systematic attempt to proactively contact subscribers outside of the initial few months of service, although many carriers have a win-back group that tries to talk subscribers out of leaving after the subscriber calls to disconnect. Some carriers have begun to look at their churn data, typically examining a small number of variables and searching for dependencies using traditional statistical models.

The goal of our research is to evaluate the benefits of predicting churn using techniques from statistical machine learning. We designed models that predict the probability of a subscriber churning within a short time window, and we evaluated how well these predictions could be used for decision making by estimating potential cost savings to the wireless carrier under a variety of assumptions concerning subscriber behavior.

II. FRAMEWORK

Fig. 1 shows a framework for churn prediction and profitability maximization. Data from a subscriber—on which we elaborate in the next section—is fed into four components that estimate

1) the likelihood that the subscriber will churn;
2) the reason for subscriber dissatisfaction given the likelihood of churn;

Manuscript received February 7, 2000; revised January 15, 2000. This work was supported in part by NSF KDI under Award 9873492.

M. C. Mozer and D. B. Grimes are with the Department of Computer Science, University of Colorado, Boulder, CO 80309-0430 USA, and are also with Athene Software, Boulder, CO 80302 USA (e-mail: mozer@cs.colorado.edu).

R. Wolniewicz, E. Johnson, and H. Kaushansky are with Athene Software, Boulder, CO 80302 USA.

Publisher Item Identifier S 1045-9227(00)04972-9.

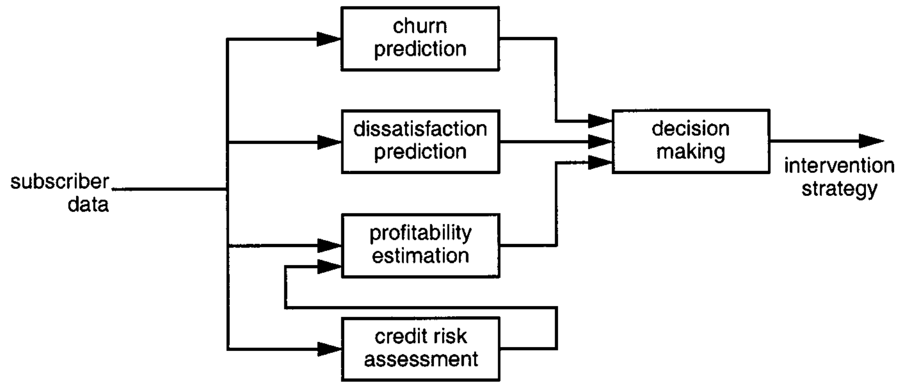

Fig. 1. Framework for churn prediction and profitability maximization

3) the profitability (expected monthly revenue) of the subscriber;
4) the subscriber's credit risk, which influences profitability.

The profitability determines how valuable the subscriber is to the carrier and, hence, influences how much the carrier should be willing to spend to retain the subscriber. Based on the predicted quantities, a decision making component determines an *intervention strategy*—whether a subscriber should be contacted, and if so, what incentives should be offered to appease them and at what cost. We adopt a decision-theoretic approach that aims to maximize the expected profit to the carrier.

In the present work, we focus on churn prediction and utilize simple measures of subscriber profitability and credit risk. However, current modeling efforts are directed at more intelligent models of profitability and credit risk. Additionally, we have found it valuable to model the likelihood of success of an intervention for a particular subscriber in order to assess the benefit of intervention.

III. Data Set

The subscriber data used for our experiments was provided by a major wireless carrier. The carrier does not want to be identified as churn rates are confidential. The carrier provided a data base of 46 744 primarily business subscribers, all of whom had multiple services. (Each service corresponds to a cellular telephone or to some other service, such as voice messaging or beeper capability.) All subscribers were from the same region of the United States: about 20% in major metropolitan areas and 80% more geographically distributed. The total revenue for all subscribers in the data base was $14 million in October 1998. The average revenue per subscriber was $234. We focused on multiservice subscribers because they provide significantly more revenue than do typical single-service subscribers.

When subscribers are on extended contracts, churn prediction is relatively easy. It seldom occurs during the contract period and often occurs when the contract comes to an end. Consequently, all subscribers in our database were month-to-month, requiring the use of more subtle features than contract termination date to anticipate churn.

The subscriber data was extracted from the time interval October through December 1998. Based on these data, the task was to predict whether a subscriber would churn in January *or* February 1999. The carrier provided their internal definition of churn, which was based on the closing of all services held by a subscriber. From this definition, 2876 of the subscribers active in October through December churned—6.2% of the database. Because the prediction window is over two months, the monthly churn rate is 3.1%.

TABLE I
FACTORS INFLUENCING SUBSCRIBER SATISFACTION

Factor	Importance	Nature of data required for prediction
call quality	21%	network
pricing options	18%	market, billing
corporate capability	17%	market, customer service
customer service	17%	customer service
credibility / customer communications	10%	market, customer service
roaming / coverage	7%	network
handset	4%	application
billing	3%	billing
cost of roaming	3%	market, billing

IV. Input Features

Ultimately, churn occurs because subscribers are dissatisfied with the price or quality of service, usually as compared with a competing carrier. The main reasons for subscriber dissatisfaction vary by region and over time. Table I lists important factors that influence subscriber satisfaction, as well as the relative importance of the factors [6]. In the third column, we list the type of information required for determining whether a particular factor is likely to be influencing a subscriber. We categorize the types of information as follows.

Network: Call detail records (date, time, duration, and location of all calls), dropped calls (calls lost due to lack of coverage or available bandwidth), and quality of service data (interference, poor coverage);

Billing: Financial information appearing on a subscriber's bill (monthly fee, additional charges for roaming, and additional minutes beyond monthly prepaid limit);

Customer Service: Calls to the customer service department and their resolutions;

Application for Service: Information from the initial application for service, including contract details, rate plan, handset type, and credit report;

Market: Details of rate plans offered by carrier and its competitors, recent entry of competitors into market, advertising campaigns, etc.;

Demographics: Geographic and population data of a given region.

A subset of these information sources were used in the present study. Most notably, we did not utilize market information because the study was conducted over a fairly short time interval during which the market did not change significantly. More important, the market forces were fairly uniform in the various geographic regions from which our subscribers were selected. In addition, we were unable to obtain information about the subscribers' equipment (age and type of handset used).

The information sources listed above were distributed over three distinct databases maintained by the carrier. The databases contained thousands of fields, from which we identified 134 variables associated with each subscriber that we conjectured might be linked to churn. The variables included

- subscriber location;
- credit classification;
- customer classification (e.g., corporate versus retail);
- number of active services of various types;
- beginning and termination dates of various services;
- avenue through which services were activated;
- monthly charges and usage;
- number, dates, and nature of customer service calls;
- number of calls made,
- number of abnormally terminated calls.

V. Data Representation

As all statisticians and artificial intelligence researchers appreciate, representation is key. A significant portion of our effort involved working with domain experts in the wireless telecommunications industry to develop a representation of the data that highlights and makes explicit those features that—in the expert's judgement—were highly related to churn. To evaluate the benefit of carefully constructing the representation, we performed studies using both *naive* and a *sophisticated* representations.

The naive representation mapped the 134 variables to a vector of 148 elements in a straightforward manner. Numerical variables, such as the length of time a subscriber had been with the carrier, were translated to an element of the representational vector that was linearly related to the variable value. We imposed lower and upper limits on the variables to suppress irrelevant variation and to not mask relevant variation by too large a dynamic range; vector elements were restricted to lie between -4 and $+4$ standard deviations of the variable. One-of-n discrete variables, such as credit classification, were translated into an n-dimensional subvector with one nonzero element.

The sophisticated representation incorporated the domain knowledge of our experts to produce a 73-element vector encoding attributes of the subscriber. This representation collapsed across some of the variables that, in the judgement of the experts, could be lumped together (e.g., different types of calls to the customer service department) and expanded on others (e.g., translating the scalar length of time with carrier to a multidimensional basis-function representation, where the receptive-field centers of the basis functions were suggested by the domain experts) and performed transformations of other variables (e.g., ratios of two variables or time-series regression parameters).

VI. Predictors

The task is to predict the probability of churn from the vector encoding attributes of the subscriber. We compared the churn-prediction performance of three classes of models:

1) logit regression;
2) C5.0 decision tree (a commercial software product based on the C4.5 algorithm of [6]);
3) nonlinear neural networks with a single hidden layer and weight decay [1].

The neural network model class was parameterized by the number of units in the hidden layer and the weight decay coefficient. We originally anticipated that we would require some model selection procedure, but it turned out that the results were remarkably insensitive to the choice of the two neural network parameters; weight decay up to a point seemed to have little effect, and beyond that point, it was harmful and varying the number of hidden units from 5–40 yielded nearly identical performance. We likely were not in a situation where overfitting was an issue due to the large quantity of data available; hence, increasing the model complexity (either by increasing the number of hidden units or decreasing weight decay) had little cost.

Rather than selecting a single neural network model, we averaged the predictions of an ensemble of models that varied in the two model parameters. The average was uniformly weighted.

For the decision tree and the neural network, we also examined the use of boosting. Adaboost [3] was used for the decision tree, and a variant of Adaboost that is appropriate when a classifier produces confidence ratings [8], [9] was used for the neural network. For neural net boosting, we used a network with ten hidden units and no weight decay.

VII. Methodology

We constructed predictors by selecting a model class (logit regression, decision tree, or neural network), a subscriber representation (naive and sophisticated), and a model combination technique (none, averaging, or boosting). For each predictor, we performed a ten-fold cross validation study, utilizing the same splits across predictors. In each split of the data, the ratio of churn to no-churn examples in the training and validation sets was the same as in the overall data set.

For the neural net models, the input variables were centered by subtracting the means and scaled by dividing by their standard deviation. Input values were restricted to lie in the range $[-4, +4]$. Networks were trained until they reached a local minimum in error.

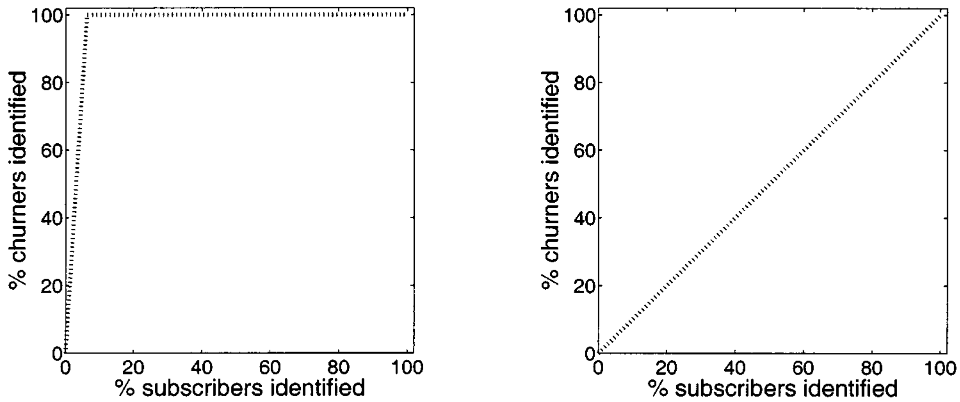

Fig. 2. (a) Lift curve indicating perfect discrimination of churners from nonchurners. (b) Lift curve indicating no discrimination of churners from nonchurners.

VIII. Results and Discussion

A. Churn Prediction

For each predictor, we obtain an estimate of the probability of churn for each subscriber in the data set by merging the test sets from the ten data splits. Because decision making ultimately requires a "churn" or "no churn" prediction, the continuous probability measure must be thresholded to obtain a discrete predicted outcome. In the telecommunications industry, the outcome is often expressed using a *lift curve*. The lift curve is related to the ROC curve of signal detection theory [4] and the precision-recall curve in the information retrieval literature.

For a given threshold on the probability of churn, we determine two quantities: 1) the fraction of all subscribers having churn probability above the threshold and 2) the fraction of all churners having churn probability above the threshold. The lift curve plots one quantity against the other. Fig. 2(a) shows a lift curve indicating perfect discrimination of churners from nonchurners. Churners have higher predicted probability than nonchurners; consequently, as the decision threshold is lowered from 1.0, only churners are identified. Because the churn rate is 6.2%, when the decision threshold is lowered such that 6.2% of the subscribers are above the threshold, 100% of the churners have been identified, and the curve reaches asymptote. Fig. 2(b) shows a lift curve indicating no discrimination of churners from nonchurners. This curve is what one would expect if the churn probabilities are random. The proportion of churners identified grows at the same rate as the proportion of subscribers identified. Thus, the more bowed the curve is to the upper-left corner of the graph, the better the predictor is at discriminating churners from nonchurners. The reason why the lift curve is relevant to the telecommunications industry is that customer service centers have a fixed staff that is able to contact a fixed fraction of the subscriber base in a given month, say, 10%. The wireless carriers are interested in estimating what fraction of churners they will catch if 10% of the subscriber base were contacted. Based on Fig. 2(a), 100% of the churners will be caught; based on Fig. 2(b), 10% of the churners will be caught.

Fig. 3 shows the lift curves for neural network and logit regression predictors, using the naive and sophisticated representations. As the figure indicates, discriminability is clearly higher for the sophisticated representation than for the naive representation. Further, for the sophisticated representation, at least, the nonlinear neural net outperforms the logit regression. It appears that the neural net can better exploit nonlinear structure in the sophisticated representation than in the naive representation, perhaps due to the basis-function representation of key variables.

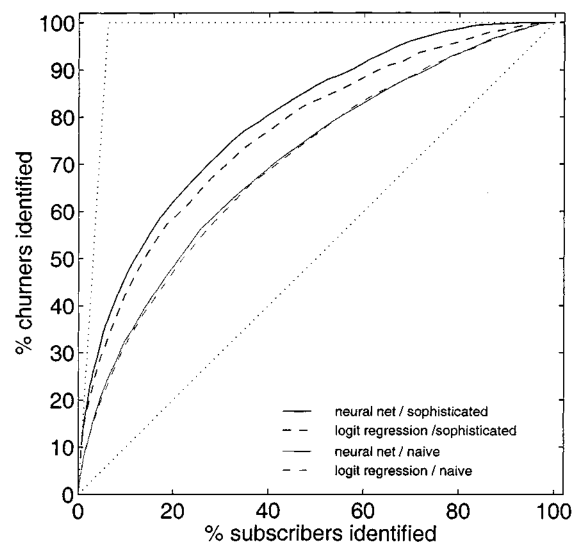

Fig. 3. Test-set performance for four predictors. The predictors are based on different models (logit regression, neural network) and input representations (naive, sophisticated).

Fig. 4 shows the lift curves for neural network and decision tree predictors both with and without boosting. For this data set, the neural net appears to outperform the decision tree, and boosting seems to help in at least the upper portion of the lift curve.

Although the various predictors appear to yield similar lift curves, they produce large differences in estimated cost savings. We describe how we estimate cost savings next.

B. Decision Making

Based on a subscriber's predicted churn probability, we must decide whether to offer the subscriber some *incentive* to remain with the carrier, which will presumably reduce the likelihood of churn. The incentive will be offered to any subscriber whose

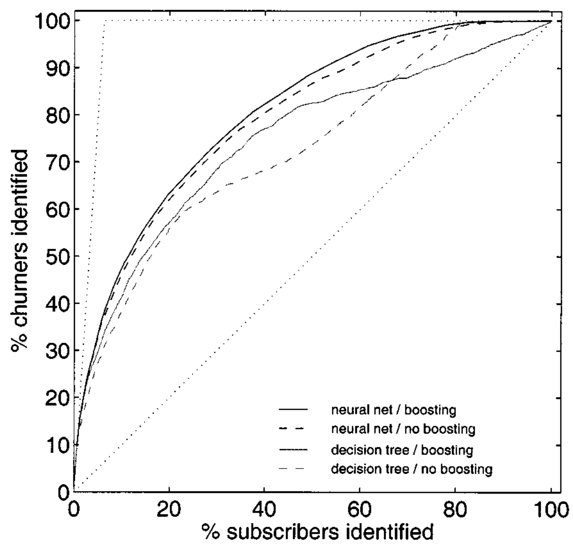

Fig. 4. Test-set performance for four predictors. The predictors are based on different models (decision tree, neural network) and whether or not boosting was applied.

churn probability is above a certain threshold. The threshold will be selected to maximize the expected cost savings to the carrier; we will refer to this as the *optimal decision-making policy*.

The cost savings will depend not only on the discriminative ability of the predictor but also on 1) the cost to the carrier of providing the incentive, which is denoted C_i (the cost to the carrier may be much lower than the value to the subscriber, e.g., when airtime is offered); 2) the time horizon over which the incentive has an effect on the subscriber's behavior; 3) the reduction in probability that the subscriber will leave within the time horizon as a result of the incentive P_i; and 4) the lost-revenue cost that results when a subscriber churns C_l.

We assume a time horizon of six months. We also assume that the lost revenue as a result of churn is the average subscriber bill over the time horizon, along with a fixed cost of $500 to acquire a replacement subscriber. (This acquisition cost is higher than the typical cost we stated earlier because subscribers in this database are high valued and must often be replaced with multiple low-value subscribers to achieve the same revenue.) To estimate cost savings, the parameters C_i, P_i, and C_l are combined with four statistics obtained from a predictor:

$N(pL, aL)$: number of subscribers who are predicted to leave (churn) and who actually leave, barring intervention;

$N(pS, aL)$: number of subscribers who are predicted to stay (nonchurn) and who actually leave, barring intervention;

$N(pL, aS)$: number of subscribers who are predicted to leave and who actually stay;

$N(pS, aS)$: number of subscribers who are predicted to stay and who actually stay.

Given these statistics, the net cost to the carrier of performing no intervention is

$$net\ (no\ intervention) = [N(pL,\, aL) + N(pS,\, aL)]C_l.$$

This equation says that whether or not churn is predicted, the subscriber will leave, and the cost per subscriber will be C_l. The net cost of providing an incentive to all subscribers whom are predicted to churn can also be estimated as

$$\begin{aligned} net\ (incentive) = &[N(pL,\, aL) + N(pL,\, aS)]C_i \\ &+ [P_i N(pL,\, aL) + N(pS,\, aL)]C_l. \end{aligned}$$

This equation says that the cost of offering the incentive C_i is incurred for all subscribers for who are predicted to churn, but the lost revenue cost will decrease by a fraction P_i for the subscribers who are correctly predicted to churn. The savings to the carrier as a result of offering incentives based on the churn predictor is then

$$\begin{aligned} &savings\ per\ churnable\ subscriber \\ &\quad = [net\ (no\ intervention)\ -\ net(incentive)] \\ &\quad\quad \cdot /[N(pL,\, aL) + N(pS,\, aL)]. \end{aligned}$$

The contour plots in Fig. 5 show expected savings per churnable subscriber for a range of values of C_i, P_i, and C_l, based on the optimal policy and the sophisticated neural-net predictor. Each plot assumes a different subscriber retention rate ($= 1 - P_i$), given intervention. The "25% retention rate" graph supposes that 25% of the churning subscribers who are offered an incentive will decide to remain with the carrier over the time horizon of six months. For each plot, the cost of intervention (C_i) is varied along the x-axis, and the average monthly bill is varied along the y-axis. (The average monthly bill is converted to lost revenue C_l by computing the total bill within the time horizon and adding the subscriber acquisition cost.) The shading of a region in the plot indicates the expected savings, assuming the specified retention rate is achieved by offering the incentive. The grey-level bar to the right of each plot translates the shading into dollar savings per subscriber who will churn, barring intervention. Because the cost of the incentive is factored into the savings estimate, the estimate is actually the net return to the carrier.

The white region in the lower right portion of each graph is the region in which no cost savings will be obtained. As the graphs clearly show, if the cost of the incentive needed to achieve a certain retention rate is low and the cost of lost revenue is high, significant per-subscriber savings can be obtained.

As one might suspect in examining the plots, what is important for determining per-subscriber savings is the ratio of the incentive cost to the average monthly bill. The plots clearly show that for a wide range of assumptions concerning the average monthly bill, incentive cost, and retention rate, a significant cost savings is realized.

The plots assume that all subscribers identified by the predictor can be contacted and offered the incentive. If only some fraction F of all subscribers are contacted, then the estimated savings indicated by the plot should be multiplied by F.

To pin down a likely scenario, it is reasonable to assume that 50% of subscribers can be contacted, 35% of whom will be retained by offering an incentive that costs the carrier $75, and in our database, the average monthly bill is $234. Under this scenario, the expected savings—above and beyond recovering the incentive cost—to the carrier is $93, based on the sophisticated neural net predictor. In contrast, the expected savings is only

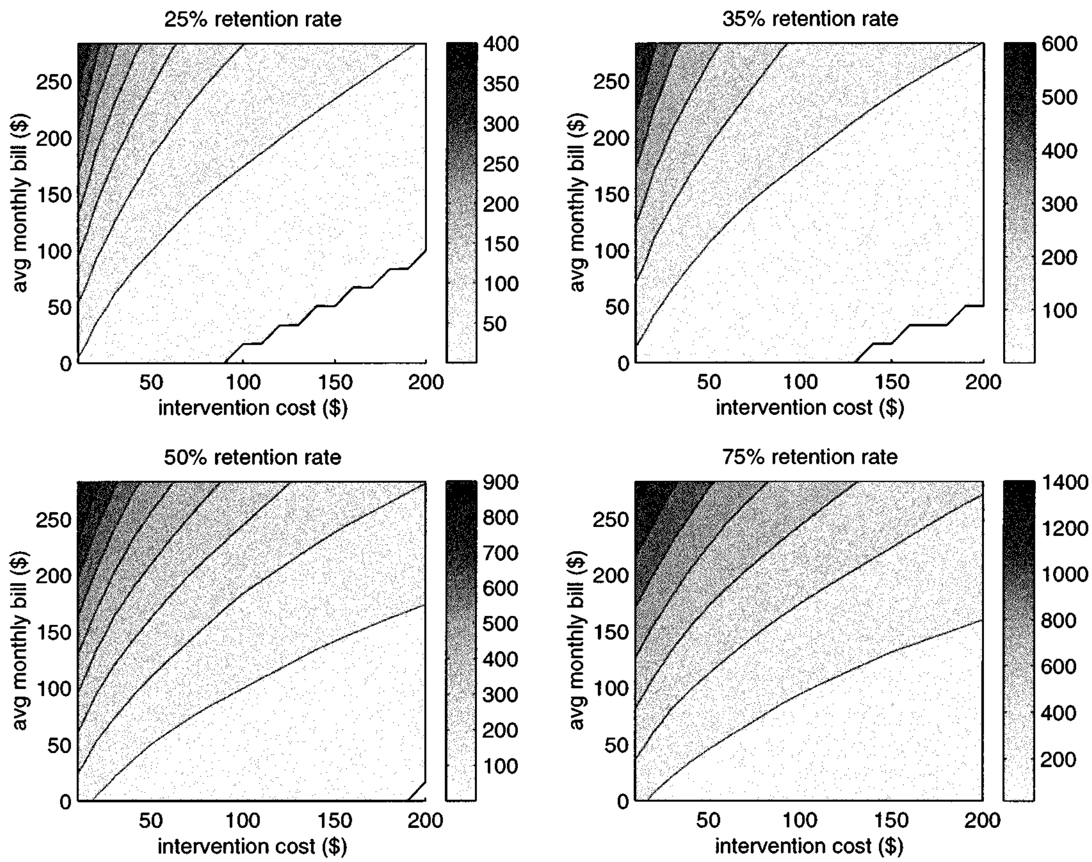

Fig. 5. Expected savings to the carrier per churnable subscriber under a variety of assumptions concerning intervention cost, average monthly bill of subscriber, and retention rate that will be achieved by offering an incentive to a churnable subscriber.

$47, based on the naive neural net predictor, and $81, based on the sophisticated logit regression model. As we originally conjectured, both the nonlinearity of the neural net and the bias provided by the sophisticated representation are adding value to the predictions.

C. Real-World Testing

The results presented to this point have been from cross-validation studies in which the test set completely overlaps in time with the training set. That is, we trained a predictor to estimate the likelihood of churn in January or February 1999, based on data from a subset of subscribers, and evaluated the predictor on January/February churn prediction for the remaining subscribers. In real-world usage, however, one would train the predictor on *all* subscribers at a given point in time, say, to predict January/February churn, and then test the following months, predicting March/April churn. Because the wireless market could be highly nonstationary, one might question the validity of our results based on a single window of time.

To evaluate the quality of predictions with a test window shifted in time relative to the training window and to evaluate the predictors in the real world, the carrier who provided us with data conducted a six-week experiment. Subscribers were randomly split into *control* and *treatment* groups, and we were blind to this split. We trained a neural network ensemble (forming the ensemble by averaging outputs) on input data from a March through August 1999 time window. This ensemble was used to recommended subscribers with high churn probability for intervention. The carrier contacted the most likely churners in the treatment group but did nothing for customers in the control group. We retrained networks weekly, using a sliding time window, and repeated the procedure for six one-week periods.

Total churn over the six week period was 3.7% in the control group but only 2.2% in the treatment group. The drop in churn was 40% as a result of treatment. The cost of intervention to the carrier was $92 per subscriber of which $17 was for the incentives offered and about $75 for the call center staff and facilities. The $75 per contact internal cost is quite high[1] and was due to the fact that the call center was being operated on a small scale for this experiment and because customer care representatives spent significant time addressing the concerns and questions of subscribers who indicated a lack of satisfaction with the carrier. Nonetheless, by our decision-theoretic framework, if a 40% reduction in churn is obtained by spending $92 for intervention on accounts valued at $234 per month, the savings per churnable subscriber is $417. As a reminder, this figure is the savings *after* the intervention cost has been incorporated.

[1]Industry data suggest that this level of expense is abnormal. It should be appreciated that when typical internal costs are supplied, the total saving associated with the reported levels of detection and retention will be significantly higher.

IX. Ongoing Research

Our ongoing research involves extending our initial results in several directions. First, we have confirmed our positive results with data from several different time windows. Second, we have further tuned and augmented our sophisticated representation to obtain higher prediction accuracy, using exploratory data-analysis techniques. Third, we are applying a variety of techniques, including sensitivity analysis, committees consisting of several different types of models, and Gaussian processes [10], to further improve prediction accuracy. Fourth, we are exploring input variable selection techniques. Fifth, we have begun to explore the consequences of iterating the decision-making process and evaluating savings over an extended time period. Regardless of these current directions of research, the results presented here show the promise of data mining in the domain of wireless telecommunications. As is often the case for decision-making systems, the predictor need not be a perfect discriminator to realize significant savings.

Acknowledgment

This work greatly benefited from the suggestions and assistance of many members of the Athene Software staff, most notably D. Howlett, P. Urbas, R. Weems, C. Guerra, L. Yan, and R. Dodier.

References

[1] C. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Press, 1995.
[2] A. J. Fowlkes, A. Madan, J. Andrew, and C. Jensen, *The Effect of Churn on Value: An Industry Advisory*, 1999.
[3] Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm,” in *Proc. Thirteenth Int. Conf. Machine Learning*, L. Saitta, Ed., 1996, pp. 148–156.
[4] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.
[5] L. Luna, “Churn is epidemic,” Radio Commun. Rep., Dec. 14, 1998.
[6] J. D. Power & Associates, “1998 Residential Wireless Customer Satisfaction Survey,”, Sept. 22, 1998.
[7] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.
[8] R. E. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” Machine Learning, to be published.
[9] H. Schwenk and Y. Bengio, “Boosting neural networks,” Neural Comput., to be published.
[10] C. K. I. Williams and C. E. Rasmussen, “Gaussian processes for regression,” in *Advances in Neural Information Processing Systems VIII*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, pp. 514–520.

Michael C. Mozer received the Ph.D. degree in psychology and cognitive science from the University of California, San Diego, in 1987.

Following a postdoctoral fellowship with G. Hinton, he joined the faculty at the University of Colorado, Boulder, and is presently an Associate Professor in the Institute of Cognitive Science and the Department of Computer Science. He also serves as the Director of Machine Learning Technologies for Sensory, Inc., Boulder, and as Chief Scientist for Athene Software, Boulder. His research interests include computational models of human perception, attention, and awareness, and applications of machine learning techniques for home automation. He has over 70 publications in neural networks.

Dr. Mozer was the General Chair of the Neural Information Processing Systems (NIPS) conference in 1996 and serves on the NIPS board of directors. He received the 1990 National Science Foundation Presidential Young Investigator Award.

Richard Wolniewicz received the B.S. degree in engineering physics in 1988, the M.S. degree in computer science in 1991, and the Ph.D. degree in computer science, database systems in 1994, all from the University of Colorado, Boulder.

He joined Athene Software, Boulder, in August 1998 as the Vice President of engineering. Prior to joining the company, he worked on projects to implement client/server applications and object-oriented database access libraries at a number of Boulder companies, including XVT Software and HeurisTec Corp. He has over ten years of commercial software development experience.

David B. Grimes (S’99) is an undergraduate in computer science at the University of Colorado at Boulder. He is planning on entering a Ph.D. program in computer science in the fall of 2000.

His research interests include statistical machine learning, probabilistic reasoning in noisy environments, Bayesian inference, and data mining. For the past three years, he has worked at the National Oceanic and Atmospheric Administration, researching a system for global satellite image transmission using wavelet compression. Currently, he is working on an honors thesis, researching connections between symbolic and subsymbolic processing in graph matching and anagram problem solving.

Mr. Grimes is student member of the ACM.

Eric Johnson received dual degrees in biology and psychology from Claremont College, Claremont, CA, in 1981.

He has served as President and CEO of Athene Software, Boulder, CO, since the company’s inception. Prior to founding Athene, he served as President, Chief Executive Officer, and Chairman of the board of directors of Coral Systems, Inc. LightBridge, Inc. acquired Coral Systems in November 1997. Prior to founding Coral Systems, he served as Director of marketing and strategic planning for ACI Telecom (a subsidiary of US WEST, which was later sold to Tandem Computers), Director of business development for PacTel Cellular (now AirTouch Cellular), President and CEO of NCAS, Inc., a company he founded that provided a wide range of services to the legal industry, and as a Genetic Technician for the Salk Institute, La Jolla, CA. He is a noted speaker at various telecommunication technology conferences and financial symposiums.

Howard Kaushansky received the B.Sc. degree from the University of California, Los Angeles, in 1980 and the J.D. degree from the University of Southern California, Los Angeles, in 1983, where he graduated Order of the Coif.

He joined Athene Software, Boulder, CO, in November 1997 as the Vice President of business development. Prior to joining Athene, he held the position of Vice President of strategic planning and general counsel for Coral Systems, Inc., a software development company focused on telecommunications subscriber fraud and retention management. He has more than eight years experience in the communications industry, with an emphasis on structuring successful business relationships with infrastructure and service providers. He holds and has filed several patents in the communications industry.