Bayesian Network Approach to Predict Mobile Churn Motivations: Emphasis on General Bayesian Network, Markov Blanket, and What-If Simulation

Kun Chang Lee[1] and Nam Yong Jo[2]

[1] Professor of MIS and WCU Professor of Creativity Science
SKK Business School and Department of Interaction Science
Sungkyunkwan University
Seoul 110-745, Republic of Korea
kunchanglee@gmail.com, leekc@skku.edu
[2] PhD Candidate
SKK Business School
Sungkyunkwan University
Seoul 110-745, Republic of Korea
namyong.jo@gmail.com

Abstract. As mobile telecommunication service becomes indispensable to our daily life, predicting the reasons of mobile churn seems essential from the perspective of mobile service providers. Previous studies have been focused on mobile churn prediction itself, not churn motivations which can play as a good indicator to forecasting real churn. Therefore, main focus of this study is placed on predicting mobile churn motivations, instead of mobile churn prediction. We propose BN approach to predict mobile churn motivation, adopting three types of BN models such as Naïve BN (NBN), Tree Augmented NBN (TAN), and General BN (GBN). To prove its validity in predicting mobile churn motivations, benchmarking classifiers were adopted and their performance was compared with BN classifiers. Through analyzing the empirical results, we found three advantages of GBN-(1) GBN performance is competitive compared with other benchmarking classifiers, (2) Markov Blanket (MB) variables are considerably small in number and make it handy for decision makers, and (3) what-if simulation is possible, which is not possible in other benchmarking classifiers. Practical implications of empirical results were addressed.

Keywords: Mobile Churn Prediction, Churn Motivation, Bayesian Network, Markov Blanket, What-If Simulation.

1 Introduction

The mobile telecommunication industry is a cut-throat world. Mobile customers seek out emerging devices and services. The annual churn rate ranges from 20% to 40% for most global mobile telecommunications service companies [9]. Reducing churn is important because acquiring new customers is more expensive than retaining existing

T.-h. Kim et al. (Eds.): FGIT 2010, LNCS 6485, pp. 304–313, 2010.

customers [17]. In order to manage customer churn to increase profitability, companies need to predict churn behavior, however this problem not yet well understood [1].

Thus, to address churn, we propose a new approach based on Bayesian network (BN). First, we present a BN structured learning algorithm to uncover the underlying motivations of churn. Second, we propose the concept of a Markov Blanket (MB) as a robust feature (variable) selection method to make our model more parsimonious. Forth, benchmarking test is performed to compare BNs with other popular classifiers. Finally, we identify the underlying churn motivations and integrate the decision-making procedure. Insight on motivations of customers churn is gained by interpreting the probabilities in these causal prediction models.

The paper is organized as follows: Section 2 contains a brief overview of related works and introduces the BN prediction models. Section 3 reviews the overall framework, a detailed description of the experiments and the results. Finally, the major contributions of the paper, managerial implications, and future research directions are discussed in Section 4.

2 Theoretical Background

2.1 Mobile Churn

Churn occurs when a customer terminates the use of a service from the service provider either voluntarily or involuntarily. In the telecommunication market, churn can be measured as the cancelation rate in a certain period of time [17]. Studies on predicting churn rate in the telecommunication market often have used Logit and Probit statistics methods. Recently, many artificial intelligence methods such as neural networking (NN) and decision trees (DT) have been used to study churn. Contractual variables and phone call-related variables were used to evaluate the explanatory variables of churn [18].

Most researches in mobile domain can be categorized into two kinds of styles. The first category focuses on causal variables to determine customer satisfaction or usage pattern through logistic regression (Logit) or structural equation models (SEM) using survey data [1] [3]. The second category relates to churn prediction, especially to determine better classifiers which have better classification accuracy [11] [13]. Thus, prior research reveals that the primary focus of previous churn prediction models has been limited to maximizing predictability, with little attention given to the issues of the motivation of customers' churn.

2.2 Bayesian Network

A Bayesian network (BN) can be defined as a directed acyclic graph (DAG) which has a probabilistic causal relationship and direction. Bayesian networking has been used effectively as a tool to support decision making in finance and marketing. [16] used BN to develop an early warning system for bankruptcy, and [7] utilized BN to increase the quality of loan assessment.

From the structural point of view, the most general type of BNs is Naïve Bayesian Network (NBN), Tree Augmented Naïve Bayes (TAN) and General Bayesian network (GBN). NBN shows a shape in which a class node is linked with all of the children nodes which are explanatory variables for the target variable. All of the explanatory attributes are dependent on the class node but independent on each other. In NBN, the class node is a special variable distinguished from other nodes, and the model assumes too much independence between variables. Therefore, it has been regarded as too rigid and not a proper reflection of reality, leading to the introduction of TAN by Friedman et al. [6] to compensate for those weaknesses. TAN expands NBN into a tree shape. Unlike the first two BNs, the GBN does not allow any differences between class nodes and general attribute nodes to express inter-dependency in the Bayesian network [2]. Therefore, class nodes can also have parent nodes, and the causal relationships or inter-dependencies among various variables in a given decision-making issue can be most naturally expressed. Rather than searching for a fully unrestricted general BN structure, as in an unsupervised Bayesian learning algorithm, the GBN classifier can seek to optimize classification performance based primarily on the probabilistically salient Markov blanket nodes.

In this paper, we use all three types of BN structure to compare performance as well as to reveal the underlying structure of churn motivations. In addition, only variables in the Markov blanket are used for further analysis to see if performance of GBNs is still competing. The concept of the Markov blanket (MB) was first introduced by Pearl [14] but has recently received renewed attention in the areas of Bayesian learning [4] and features selection [10]. The probabilistic nature of the Markov blanket can be explained using the concept of *d-separation* (*direction dependent separation*) [14], a graphical criterion related to the blocking of information flow among variables. In a faithful Bayesian network, *d-separation* captures all of the conditional independence relationships encoded in the network [14]. If all variables in MB of a node are instantiated, then the node is *d-separated* from the rest of the network. In other words, $MB(T)$ is a minimal feature subset required to predict T, which graphically corresponds to a set neighborhood of T: its parents, its children and the other parents of its children.

3 Experiments

3.1 Data and Variables

The data in this paper were donated by a major mobile telecommunication company in South Korea. This is originally consisted of 14 variables and 5,000 records that were randomly sampled from anonymous churned customers. After removing records containing a missing value 4,922 records are remained for experiments in this research.

Table 1 summarizes the variables used in this study. Customers' age (Age), makers of the mobile devices (DeviceMaker), customers' service grades provided by companies (ServiceGrade) and the way customers pay the bill (Paymentway) are variables originally donated by categorized. Continuous variables reflect the number of calls to the contact center in a previous month (CallsInaMon), average revenue per user in a previous month (ARPUInaMon), average revenue per user over previous three months (ARPUIn3MonAve), the frequencies of voice calls in a previous month (VoiceFreqInaMon) and the minutes of voice usage in a previous month (VoiceMinuInamon). These variables were meaningfully transformed into categorical variables using a supervised discrete method supported by WEKA[1]. We inserted two more additional

[1] Available at `http://www.cs.waikato.ac.nz/ml/weka/`

continuous variables to characterize customers' loyalty such as months of usage (totMonofUsg) and the duration ratio of usage after changing to a new device out of the total months of usage (AfterdevchgP). Finally, we purposely categorized customers' churn motivation (ChurnMotive) which is target variable into major five variables by integrating values of small frequencies into one value, 'others'. In this way, the 12 variables were prepared for our experiments.

Table 1. Tweleve variables and available values for learning classifiers

Variable	Available values (number of values)
Age	10under, 10to12, 13to15, 16to19, 20to24, 25to29, 30to34, 35to39, 40to44, 45to49, 50to54, 55to59, 60over (13)
DeviceMaker	lge, motorola, pantech, samsung, skteletech, others (6)
ServiceGrade	BRONZE, GOLD, SILVER, VIP (4)
PaymentWay	AUTOBANK, CREDITCARD, JIRO (3)
CallsInaMon	-0.5, 0.5-2.5, 2.5- (3)
ARPUInaMon	-21, 21-3537, 3537-3599.5, 3599.5-9501, 9601-14810, 14810-33173.5, 33173.5-(7)
ARPUIn3MonAve	-2675, 2675-3567.5, 3567.5-4925.5, 4925.5-9444.5, 9444.5-15780.5, 15780.5-34978.5, 34978.5- (5)
VoiceFreqInaMon	-0.5, 0.5-85.5-, 85.5- (3)
VocieMinuInaMon	-7.5, 7.5-6854, 6854- (3)
totMonofUsg *	*-37.5, 37.5- (2)*
AfterdevchgP *	*-0.905, 0.905-, (2)*
ChurnMotive **	***failtopay, noneeds, numtrans, others(eg. Burden of high bill, stop to use of specific service (5)***

* manipulated by author, all other variables were left intact as donated.
** class node (target variable).

3.2 BN Classifiers

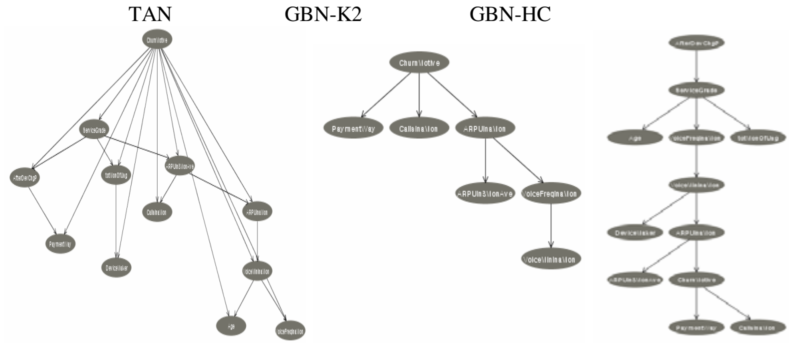
3.2.1 Structure Learning

We used WEKA [8], an open source data mining tool for various kinds of data mining, for BN constructs and experimental performance. The 11 variables (Table 2) were used to determine which networks had a target node of 'Churn motivation.' The structure of the GBN was learned using two search algorithms, K2 [5] and Hill Climber, with the maximum number of parent nodes limited to one. A BDeu scoring metric was used, a refinement of the K2 metric. The MDL scoring metric was used for Hill Climber (HC) [12]. For constructing NBN and TAN [6], the default setting in WEKA was used.

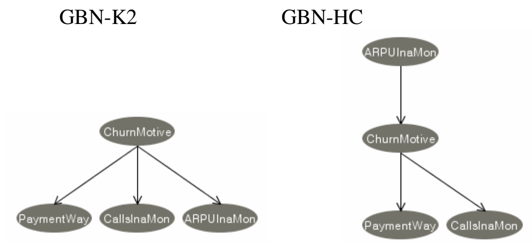
Table 2. Two types of variables

	Structure	NBN	TAN	GBN-K2	GBN-HC
Num. of nodes	Full variables	12(11)	12(21)	7(6)	12(11)
(Num. of Arc)	MB variables	-	-	4(3)	4(3)

Our experiments to create BN structures involved two types of variables, one for full variables and the other for MB variables. The later is purposed that only variables included in MB were used to create a structure, so the accuracy of classification was stable. Table 2 illustrates these types. The final analysis was conducted on 6 structures, which are illustrated in Figure 1. We purposely omitted NBN, because its shape is straightforward such that all the children nodes are directly linked with class node.



(a) Full variables



(b) Markov Blanket (MB) variables

Fig. 1. BN structures using two types of variables-full variables and MB variables

3.2.2 Results

(1) Prediction accuracy

For the sake of clear understanding of the experiment performance, we address the results from BN methods, and then the benchmarking classifiers' performance will be compared with BN performance.

Experiments with BN classifiers

First of all, let us state the results from experiments using full variables. Table 3(a) lists the performance of BN classifiers. Experiments were conducted with ten repetitions of ten-fold cross-validation on classification of mobile churn motivation. Prediction accuracy was measured in percent and figures in parenthesis indicate standard deviation. The best result is underlined in Table 3(a). To obtain rigor for each BN classifier's performance, each BN performance was compared with other BN classifiers' performance by using a corrected resample t-test at the 5% significance level based on the 10 x 10 cross-validation results. The results in Table 3(b) reassure that (1) GBN-HC and TAN show the statistically same performance, and (2) both GBN-K2 and NBN turn out to be equivalent statistically. In addition, comparing performance of BN classifiers, we came to understand that GBN-HC and TAN outperform GBN-K2 and NBN.

By the way, as stated previously, number of MB variables is considerably small compared with number of full variables. Furthermore, performance using MB variables is not bad in comparison with performance using full variables. This argument is confirmed by looking at the *Full variables vs MB variables* part of Table 3(b) where GBN-K2 performance using MB variables (47.68%) is better than GBN-K2 using full variables (47.16%), and GBN-HC using full variables shows statistically better performance than GBN-HC using MB variables (48.23% > 47.68%).

Judging from the discussion so far, it is clearly concluded that performance of BN classifiers using MB variables possess great potentials- (1) complexity measured by number of variables and arcs is very low in comparison with full variables, and (2) performance is not defeated seriously by the BN classifiers using full variables.

Table 3. Prediction Accuracies of BN Classifiers and Statistical Tests for Comparison

(a) Accuracies for each BN (unit: %)

Structure	NBN	TAN	GBN-K2	GBN-HC
Full variables	46.78(1.62)	48.27(1.58)	47.16(1.83)	48.23(1.72)
MB variables	-	-	47.68(1.72)	47.68(1.73)

(b) t-test results for BN classifiers

	Paired Differences					t-value	Sig. (2-tailed)
				95% Confidence Interval of the Diff.			
	Mean	Std. Dev.	Std. Error Mean	Lower	Upper		
Full variables							
NBN - TAN**	-1.49	1.99	0.199	-1.89	-1.09	-7.49	0.00
NBN – GBN K2	-0.38	2.18	0.218	-0.82	0.05	-1.76	**0.08**
NBN - HC**	-1.46	2.40	0.240	-1.94	-0.98	-6.07	0.00
TAN - K2**	1.11	2.42	0.242	0.63	1.59	4.58	0.00
TAN – GBN HC	0.03	2.36	0.236	-0.44	0.50	0.14	**0.89**
K2 - HC**	-1.07	2.23	0.223	-1.52	-0.63	-4.82	0.00
Full variables vs MB variables							
Full GBN K2 -MB GBN K2	-0.52	1.25	0.125	-0.77	-0.28	-4.21	0.00
Full GBN HC- MB GBN HC	0.56	0.69	0.069	0.42	0.69	8.07	0.00

** $p < 0.05$, ** $p < 0.01$.*

Experiments with benchmarking classifiers

Benchmarking classifiers are necessary to verify the validity of accuracies of our proposed BN approach in Table 4. They include classical types of classifiers such as SVM (Support Vector Machine), Neural network (NN), and Decision tree. For the sake of SVM, LibSVM[2] was used. DT was C4.5 [15] and its module was adopted from *J48* of WEKA. NN used for benchmarking classifier was a *MultilayerPerceptron* module supported by WEKA. For the sake of computational rigor, 10-fold cross-validation was performed 10 times, and average performance was obtained for each classifier. To show that the GBN-based approach predicting mobile churn motivations outperforms benchmarking classifiers and other types of BNs, we performed paired-samples t-test

[2] `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

using SPSS 12.0. Table 4 summarizes prediction accuracies of benchmarking classifiers. Test results show that prediction accuracies of SVM, C4.5, TAN and GBN-HC are not different statistically from each other.

Table 4. Benchmarking Classifiers and Its Comparison with BN classifiers

(a) Performance of benchmarking classifiers (unit: %)

Classifiers	SVM	NN	J48
performance	42.82(1.61)	43.92(2.30)	47.96(1.82)

Figures in parenthesis indicate standard deviation.

(b) t-test results for benchmarking classifiers

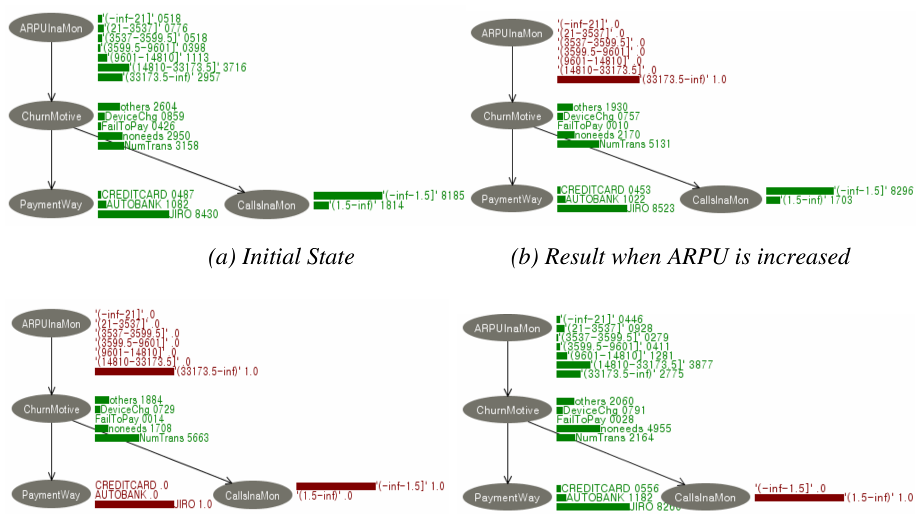
	Paired Differences					t-value	Sig. (2-tailed)
	Mean	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Diff. Lower	Upper		
SVM - NN**	3.89	2.46	0.246	3.41	4.38	15.82	0.00
SVM - J48	-0.14	2.36	0.235	-0.61	0.32	-0.61	**0.54**
SVM - NBN**	1.04	2.19	0.216	0.61	1.48	4.74	0.00
SVM - TAN	-0.45	2.39	0.239	-0.92	0.02	-1.88	**0.06**
SVM - K2**	0.66	2.42	0.242	0.18	1.14	2.71	0.01
SVM - HC	-0.42	2.33	0.233	-0.88	0.04	-1.79	**0.08**
NN - J48**	-4.04	2.85	0.285	-4.60	-3.47	-14.17	0.00
NN - NBN**	-2.85	2.86	0.286	-3.42	-2.29	-9.99	0.00
NN - TAN**	-4.34	2.83	0.283	-4.91	-3.78	-15.37	0.00
NN - K2**	-3.24	3.06	0.306	-3.85	-2.63	-10.57	0.00
NN - HC**	-4.31	2.84	0.284	-4.88	-3.74	-15.18	0.00
J48 - NBN**	1.18	2.41	0.241	0.71	1.66	4.92	0.00
J48 - TAN	-0.31	2.49	0.249	-0.80	0.187	-1.23	**0.22**
J48 - K2**	0.80	2.55	0.255	0.29	1.31	3.13	0.00
J48 - HC	-0.27	2.55	0.255	-0.78	0.23	-1.07	**0.29**

* $p < 0.05$, ** $p < 0.01$.

Then, at this moment, we have to raise two criteria by which method can be compared with each other fairly. First criterion is whether causal relationships can be induced from the method. Values of causal relationships should be measured by the characteristics of our target problem- prediction of mobile churn motivations. For the mobile telecommunication service providers investigating motivations of users to churn, such causal relationships are essential. In addition, number of churning motivations are numerous, indicating that company should spend a lot of time and efforts to pin down exact reasons of churning and amend operational and/or marketing pitfalls to prevent further churning. Second criterion is whether the number of variables can be reduced logically without loss of prediction accuracy. From this perspective, the GBN is very good at reducing the number of variables to be considered because its MB properties guarantee such possibility. This means telecommunication service providers are able to predict churn motivation more efficiently with compacter form of causal relations among MB variables. Thus, we continue to analyze practical implication with BN classifiers using MB variables in the following section.

(2) What-if analysis

The motivation of churn can be predicted by performing what-if analyses with GBN structure. Let us investigate what-if results with GBN-HC structure using MB variable. Mobile telecommunications service providers want to know how many customers are likely to transfer to competing companies, which is represented by the 'NumTrans' value in the class node 'ChurnMotive'. Figure 2 illustrates the what-if analysis results showing that customers are more likely to transfer to a competing company when (1) their ARPU of the previous month was rather high (2) they pay their bill using JIRO and (3) they call contact center very few.



(a) Initial State

(b) Result when ARPU is increased

(c) Result when payment method is JIRO and there are few calls for contact center

(d) Result when there are few calls for contact center

Fig. 2. What-if analyses using MB variables

3.3 Discussion

First of all, it should be noted that GBN classifiers can provide a set of causal relationships among relevant variables with target node, and then decision makers are able to find useful strategies to prevent undesirable churning behavior by forecasting which kinds of customers are believed to have churning motivation. From the performance perspective, GBN classifiers are showing competitive prediction accuracies compared with other BN classifiers. Therefore, we found usefulness of GBN classifiers in the problem area of predicting customers' churning motivation.

On the other hand, value of MB variables must be mentioned here. MB variables are those variables provided by only GBN classifiers. What is striking with the MB variables is that prediction accuracies by GBN with MB variables are very competitive compared with other BN classifiers using full variables. Therefore, using the MB variables can provide a number of advantages for decision makers who seek more compact decision mechanism where number of decision variables to be considered by decision makers should be kept to minimum.

Lastly, GBN classifiers are capable of uniquely providing what-if simulation functions with which decision makers can test various numbers of alternative solutions to the target problem. In case of this paper, decision makers can perform a lot of what-if simulations before deciding a final strategy to prevent customers' churning behavior which is undesirable to companies' profitability.

4 Concluding Remarks

Previous studies about mobile churn prediction have always handled churning behavior itself. However, contrary to this research trend, this study is aimed at predicting churn motivations by using BN classifiers and then comparing their performance with benchmarking classifiers. Let us summarize our contributions as follows.

First, four types of BN classifiers were considered- NBN, TAN, GBN-K2, and GBN-HC. Besides, to show validity of MB variables provided by GBN, we tested the performance by MB variables with other BN classifiers. From this attempt, usefulness of using GBN assisted by MB variables is very high, especially in the field of decision problems where a lot of decision variables should be considered before suggesting a final solution.

Second, flexible properties of what-if simulation must also be highlighted. Such what-if simulation capability is found only in BN classifiers. Therefore, BN classifiers are recommendable to be used in resolving complicated decision problems like churning motivation prediction, once their prediction accuracy is competitive.

Further study issues remain. For example, ensemble approach seems necessary to improve prediction accuracy. Another type of future study issue is to determine relevant number of variables to describe churning motivation.

Acknowledgments. This research is supported by the Ubiquitous Computing and Network (UCN) Project, Knowledge and Economy Frontier R&D Program of the Ministry of Knowledge Economy (MKE) in Korea as a result of UCN's subproject 09C1-T2-20S.

References

1. Ahn, J., Han, S., Lee, Y.: Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. Telecommunications Policy 30(10-11), 552–568 (2006)
2. Bouckaert, R.: Properties of Bayesian belief network learning algorithms. In: Proc. 10th Annual Conf. Uncertainty Artificial Intelligence (UAI), Seattle, WA, pp. 102–110 (1994)
3. Chen, P.-Y., Hitt, L.M.: Measuring switching costs and the determinants of customer retention in Internet-enabled businesses: Astudy of the online brokerage industry. Information Systems Research 13(3), 255–274 (2002)
4. Cheng, J., Greiner, R.: Comparing Bayesian network classifiers. In: Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence, pp. 101–107. Morgan Kaufmann Publishers, San Francisco (1999)

5. Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9(4), 309–347 (1992)
6. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine Learning 29(2), 131–163 (1997)
7. Gemela, J.: Financial Analysis Using Bayesian Networks. Applied Stochastic Models in Business and Industry 17, 57–67 (2001)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)
9. Kim, M.K., Jeong, D.H.: The effects of customer satisfaction and switching barriers on customer loyalty in Korean mobile telecommunication services. Telecommunications Policy 28(2), 145–159 (2004)
10. Koller, D., Sahami, M.: Toward optimal feature selection. In: Proc. 13th International Conf. Machine Learning, pp. 284–292 (1996)
11. Lemmens, A., Croux, C.: Bagging and Boosting Classification Trees to Predict Churn. Journal of Marketing Research 43(2), 276–286 (2006)
12. Madden, M.G.: On the classification performance of TAN and general Bayesian networks. Knowledge-Based Systems 22(7), 489–495 (2009)
13. Neslin, S.A., Gupta, S., Kamakura, W., Lu, J., Mason, C.: Defection detection: Measuring and understanding the predictive accuracy of customer churn models. Journal of Marketing Research 43(2), 204–211 (2006)
14. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco (1988)
15. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
16. Sarkar, S., Sriram, R.S.: Bayesian Models for Early Warning of Bank (1998)
17. Siber, R.: Combating the churn phenomenon. Telecommunications 31(10), 77–81 (1997)
18. Yan, L., Wolniewicz, R.H., Dodier, R.: Predicting customer behavior in telecommunications. IEEE Intelligent Systems 19(2), 50–58 (2004)
19. Wei, C.P., Chiu, I.T.: Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach. Expert Systems with Application 23, 103–112 (2002)