Constraint Mining in Business Intelligence: A Case Study of Customer Churn Prediction

Nittaya Kerdprasop, Phaichayon Kongchai and Kittisak Kerdprasop

Data Engineering Research Unit,
School of Computer Engineering,
Suranaree University of Technology, Thailand
nittaya@sut.ac.th, zaguraba_ii@hotmail.com, kerdpras@sut.ac.th

Abstract

In the era of digital technologies, most enterprises have collected huge amount of data in an electronic form. Business intelligence technology has emerged as a tool to support information summarization, pattern extracting, knowledge discovery, and other knowledge-related tasks. The main part of most business intelligence software is the data mining engine to analyze and report relationships that exist in the stored data. Visualization tools are created to help data analysts easily explore the induced information. For extremely large amount of data stored in the data warehouse and data marts, simply explore information and knowledge through the visualize tool is not possible. We thus propose to put more constraints in the data mining engine of the BI software. We design the framework of the proposed BI system to predict customer churn in the telecommunication industry. The logic-based implementation and performance testing results of the constraint-based pattern mining are also illustrated in this paper.

Keywords: *Constraint data mining, Pattern induction, Business intelligence, Customer churn prediction, Constraint logic programming*

1. Introduction

Business intelligence (BI) is a broad term normally used to refer to any aspect of computer-based business applications including decision support, information management, marketing automation, and intelligent data analysis [5, 10, 14]. The task of automatically extracting patterns from data related to decision making is normally done by applying statistical techniques [5]. Such methods cannot keep pace with the exponential growth of electronic data. Business analysts are gradually exploiting a faster tool of data mining techniques. Therefore, current BI software in the market contains more or less some modules of intelligent analysis based on data mining to extract useful information hidden in the enterprise databases. The extracted information from this automatic process is however tremendous in its quantity. Analysts have to post-process the analysis results by thoroughly exploring and selecting only the most informative knowledge reported to executives. Constraint data mining is thus lately proposed by several researchers [6, 13, 22, 23, 24] as a technique to alleviate the problem of superfluous knowledge.

In this paper, we propose a framework of incorporating a constraint-based pattern mining as a knowledge discovery module in the BI software. Such module can help analysts filter from large amount of knowledge the most relevant ones to their interest. The design and implementation of our pattern mining feature are based on the association rule induction [1]. Querying the induced rules can also be performed through the logic-based language based on the Prolog syntax. The mining steps of our system are constrained by preferences, which are to be identified by analyzer. The experimentation to evaluate to performance of the proposed system has been done through the analysis of customer churn.

Customer churn, also known as customer attrition or customer turnover, is the loss of existing customers to another company or service provider. Business sectors take customer churn as a serious subject because the cost of retaining current customers is much lower than acquiring the new ones [25]. Conventional statistical methods such as logistic regression analysis [17] are normally adopted to analyze and predict churning customers. In the context of business intelligence, emerging techniques from the data mining research field can also be applied to help analyzing churn customer in more various aspects than the conventional methods. Automatic data analysis by means of data mining and machine learning technologies has long been applied to the problem of churn analysis (as summarized in Table 1).

Table 1. Summary of data mining techniques applied to churn analysis and prediction

Literature	Year	Business Sector	Data Mining Techniques
[15]	1997	Finance	Rough set theory
[11]	1998	Finance	Genetic algorithm, Rough set theory, Decision tree induction, Logistic regression
[18]	2003	Telecommunication	Naïve Bayes
[28]	2005	Telecommunication	Support vector machine, Neural network, Decision tree induction, Naïve Bayes
[27]	2006	Telecommunication	Decision tree induction, Logistic regression
[17]	2006	Finance	Logistic regression
[2]	2007	Telecommunication	Decision tree induction
[20]	2009	Telecommunication	Hybrid of decision tree induction and logistic regression
[4]	2009	Finance, Telecommunication, Mass media, Retail	Decision tree induction, Logistic regression
[21]	2010	Telecommunication	Principal component analysis and clustering
[7]	2010	Finance, Supermarket	Probabilistic tree, AdaBoost
[9]	2010	Finance	Support vector machine with rule representation
[8]	2011	Finance, Retail, Telecommunication	Boosting, Bagging
[26]	2011	Telecommunication	Ant colony optimization, Support vector machine with rule representation
[19]	2011	Shipping	Rule induction
[12]	2011	Finance	Similarity-based approach

Most work on customer churn analysis aims at inducing an accurate churner/non-churner classification model. Besides high predictive accuracy of the model, comprehensibility is also an important issue as pointed out in recent work on churn prediction [12, 19, 26]. We also agree to the comprehensibility and applicability issues in customer churn analysis. The model representation should be in a form that is easy to understand by most users, not just the experts. We thus consider deliver model induction results as a set of association rules. In this paper, we also propose a framework to incorporate induced model to the decision support system. The design and implementation of constraint-based rule induction module are main contribution of this paper.

2. Pattern Analysis Framework and Mining Method

We design the inductive customer churn analysis framework (Figure 1) with the main purpose of providing early suggestion to strategic planners before customers actually leaving the company. The content management module is the part to access customer details from the stored data. Not every single detail is to be used by the inductive module, therefore the three modules (that is, content segmentation, content conversion, and DM format manager) are necessary for screening and extracting potentially useful features from the database. Selected data in an appropriate format are then sent to the knowledge management module to analyze and induce model that can characterize customers' patterns and predict future events from current situation. These induced models are considered valuable knowledge that will be finally sent back to generate actionable suggestion to the strategic planners.

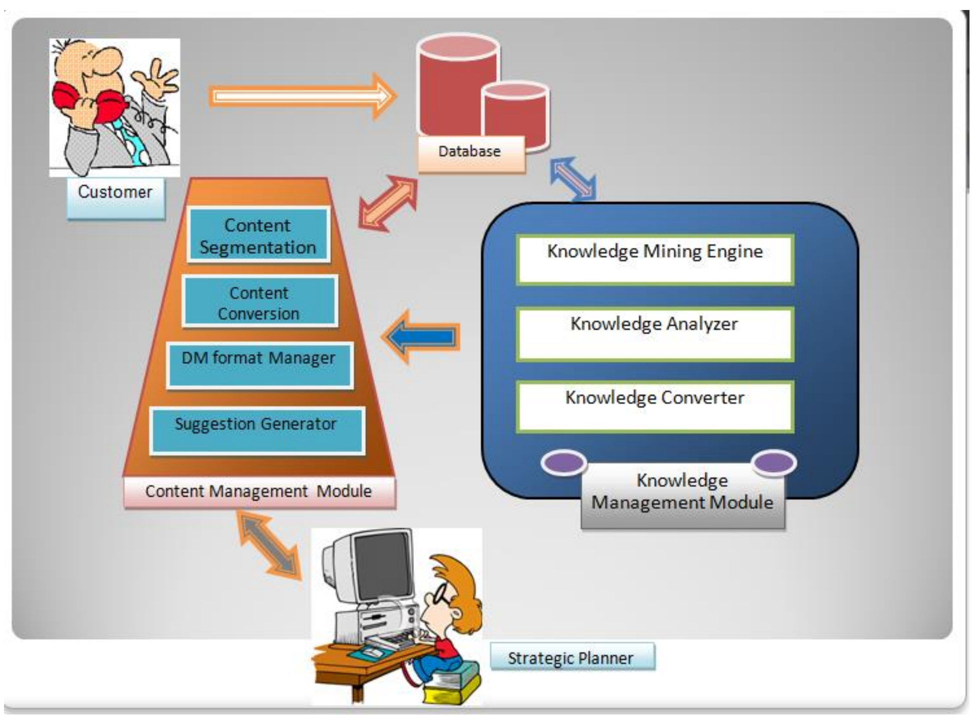

Figure 1. The pattern analysis framework to induce knowledge for supporting strategic decision

The focus of our design is the knowledge mining engine, which is the main part of the proposed business intelligence framework. The inductive algorithm in our framework is the extension of Apriori [1], which is the most well-known algorithm for association mining. We extend (in Figure 2) conventional association mining steps by considering constraints that can possibly post by analyzers or strategic planners to search for association rules that are really related to their objectives. Any irrelevant rules will automatically be removed.

```
Algorithm  Constraint-Association-Mining

//Input : Database D, Length, Subset, NotSubset, Minimum_support.
//Output :  L, frequent itemsets in D.

L_1 = find_ frequent_1itemset(D)
for(k = 2; L_{k−1} ≠ ∅; k++){
        C_k  = apriori_gen(L_{k−1}, Minimum_support);
for each transaction t ∈ D { // scan D for counts
        C_1  = subset(C_k, t)
        for each candidate c ∈ C_1 {
                c.count++
        }
C_2 = checkcondition(Length, Subset, NotSubset, C_1)
L_k = {c ∈ C_2 | c.count ≥ Minimum_support}
}  }
return ∪_k L_k
```

Figure 2. A constraint-based association mining method

3. Implementation and Running Results

We implement knowledge mining engine with the constraint-based logic programming paradigm using Eclipse 6.0 constraint system. The implementation of the constraint-based association mining is illustrated in Figure 3. On running the implemented program, we use the churn data in telecommunication industry [3, 16]. The data set contains information of 3333 customers. In the original data set, each customer record has 21 features (or variables) in which the last one is the label churn/non-churn. Details of these features are explained in Table 2. The first step of our experimentation is feature selection; the selected 12 features are state, account length, area code, international plan, voice mail plan, number vmail messages, total day calls, total eve calls, total night calls, total intl calls, number customer service calls, and churn. The other nine features are removed because of their insignificance in inducing model.

```
:-lib(ordset).
:- compile("filename.txt").   % load file.

association(R,LengthI,Subset,NotSubset,MinSup,Conf) :- data(Data), Data =_-_,
        ( count(I,2,6), fromto(Data,S0,S1,R), param(MinSup,LengthI,Subset,Conf,NotSubset) do
        ( S0=A-B,  findCL(A-B-MinSup,R-_-_, LengthI,Subset,NotSubset,Conf),
          allUnion(I,R,NewItemSet),   S1=NewItemSet-B ), !    ).

findCL(ItemSet-Items-MinSup,R-Items-MinSup,LengthI,Subset,NotSubset,Conf) :-
            ItemSet = [H|_],length(H,LenItem),
            LenItem =1 -> findSubOk(ItemSet,Items,MinSup,R,_) ;
            ( findLength(LengthI,ItemSet,ItemSet1),
            findSubset(ItemSet1,R1,Subset),  findNotSubset(R1,R2,NotSubset),
            findSubOk(R2,Items,MinSup,R,LenItem1) ), findRule(R-Items-Conf,LenItem1).

findLength(Cons,Re,R) :-   (foreach(I,Re), fromto(R,S1,S0,[]), param(Cons) do
                             length(I,LenItem),LenItem > Cons -> S1 = [ I | S0], ! ; S1=S0      ).

findSubOk(R2,Items,MinSup,R,R1) :-
             (foreach(X,R2), fromto(R,S1,S0,[]),fromto(R1,S3,S2,[]), param(Items,MinSup) do
              supOK(X,Items,MinSup,Len) -> S1=[X|S0],S3 = [Len|S2], !  ; S1=S0, S3=S2     ).

findSubset(X,X,[]).
findSubset(ItemSet,R1,[Subset|Tr]) :-  Subset = [] -> R1 = ItemSet ;
           (foreach(X,ItemSet), fromto(R,S1,S0,[]), param(Subset) do
            intersection(Subset,X,ReSub),ReSub\=[] -> S1=[X|S0],  ! ; S1=S0 ), findSubset(R,R1,Tr).

findNotSubset(X,X,[]).
findNotSubset(ItemSet,R1,[NotSubset|Tr]) :-  NotSubset = [] -> R1 = ItemSet ;
                (foreach(X,ItemSet), fromto(R,S1,S0,[]), param(NotSubset) do
                 intersection(NotSubset,X,ReSub),ReSub=[] -> S1=[X|S0],  ! ; S1=S0 ),
                findNotSubset(R,R1,Tr).

supOK(X,Items,MinSup,LenItem) :-  (foreach(I,Items), fromto(R,S1,S0,[]), param(X) do
                  (my_subset(X,I) -> S1 = [ I | S0], ! ; S1=S0)   ),
                  length(R,LenItem),   LenItem >= MinSup.

findRule([]-_-_,[]).
findRule([X|ItemSet]-Items-MinConf,[LenItem|LenItem1]) :- ItemSet = 0 -> ! ;
                 findall(Re,powerset(X,Re),PwSet),
                (foreach(ItemX,PwSet), param(X,Items,LenItem,MinConf) do
                    ( ItemX = X ; ItemX = [] -> ! ;   supOK(ItemX,Items,0,LenItemX),
                     conOk(LenItem-LenItemX-MinConf) -> createRule(ItemX,X,Re),
                     write('If '),write(ItemX), write(' '), write(LenItemX), write(' then '), write(Re),
                     write(' '), writeln(LenItem), ! ; ! ) ), findRule(ItemSet-Items-MinConf,LenItem1).
createRule([],X,X).
createRule([H|Tr],X,Result) :- delete(H,X,Re),createRule(Tr,Re,Result).

conOk(LenItem-LenItemX-MinConf) :- Re is (LenItem/LenItemX)*100,Re >= MinConf .
```

Figure 3. Implementation of the proposed constraint-association-mining algorithm

Table 2. Variable details of the customer churn data

Variable name	Data type	Description
state	discrete	Name of 50 states and District of Columbia
account length	continuous	How long account has been active
area code	continuous	
phone number	discrete	A surrogate for customer ID
international plan	discrete	Dichotomous categorical, yes or no
voice mail plan	discrete	Dichotomous categorical, yes or no
number vmail messages	continuous	Number of voice mail messages
total day minutes	continuous	Minutes customer used service during the day
total day calls	continuous	
total day charge	continuous	
total eve minutes	continuous	Minutes customer used service during evening
total eve calls	continuous	
total eve charge	continuous	
total night minutes	continuous	Minutes customer used service during the night
total night calls	continuous	
total night charge	continuous	
total intl minutes	continuous	Minutes customer used service to make international calls
total intl calls	continuous	
total intl charge	continuous	
number customer service calls	continuous	
churn	discrete	Dichotomous categorical, true or false

We then performed a series of eight experiments on the selected churn data set to induce association rules with various constraints:

Exp. 1: Rules are to be induced with thresholds: minimum support = 50 (that means there must be at least 50 records from the total of 3333 satisfying the rule's content) and minimum confidence = 80%. (The other experiments also specify the same minimum support and confidence.)

Exp. 2: Rules must contain the feature churn_False (that is, customer is non-churner).

Exp. 3: Rules must have at least three features.

Exp. 4: Rules must NOT contain the feature 'churn_False'.

Exp. 5: Rules must contain the feature 'churn_False' at the then-part of the rule.

Exp. 6: Rules must contain either the feature ‘churn_False’, or ‘churn_True’.

Exp. 7: Rules must contain both the feature ‘churn_True’ and ‘vMailPlan_no’.

Exp. 8: Rules must have at least three features, must contain both ‘churn_False’ and ‘vMailPlan_no’, must NOT contain either the feature ‘vMailMessage_0’, or ‘intlCalls_2’, and the target clause of the rules must be ‘churn_False’.

Running result of experiment 8 is shown in Figure 4. We comparatively illustrate the eight experimental results in Figure 5.

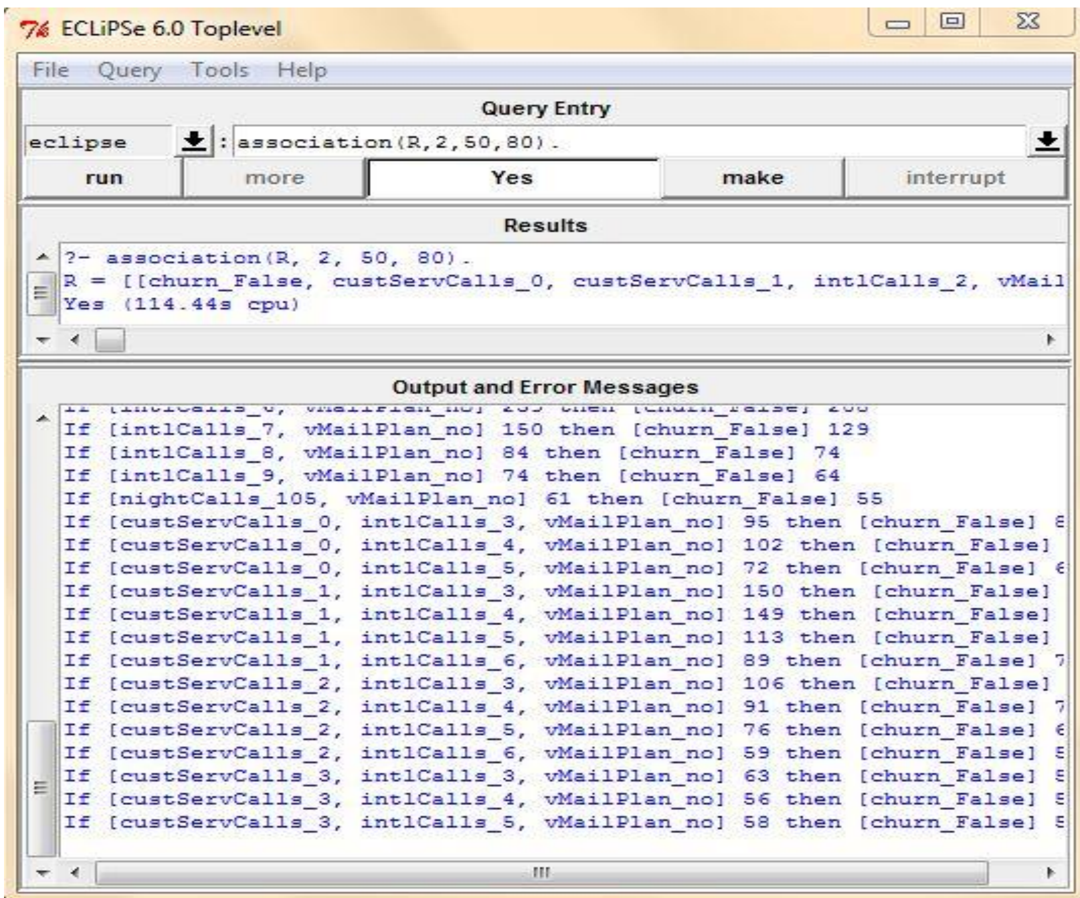

Figure 4. Running results of experiment 8: minimum support = 50 and minimum confidence = 80%. The results must contain rules that has at least three features, must contain both ‘churn_False’ and ‘vMailPlan_no’, the results must NOT contain either the feature ‘vMailMessage_0’, or ‘intlCalls_2’, and the target clause of the rules must be ‘churn_False’

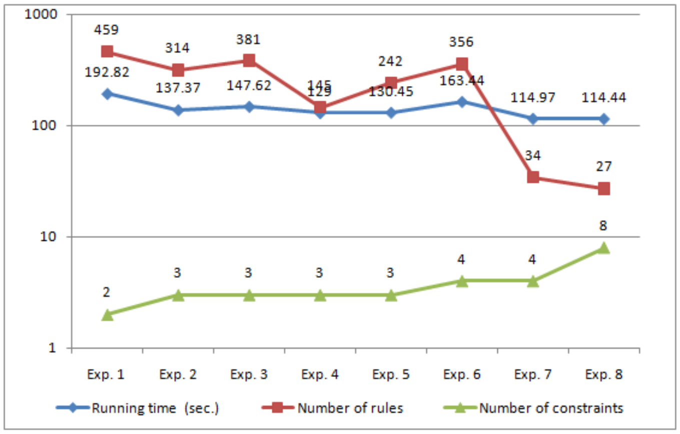

Figure 5. A comparison of computational time usage and number of rules received from varying constraints in each of the eight experiments

In Figure 5, we plot graph on a logarithmic scale for ease of comparison. Experiment 1 represents the conventional Apriori association mining in which the inherent constraints are minimum support and confidence. From experiments 2 to 8, we extend the constraints to the specification of the desired association rules such as features that must appear/not-appear in the final result, number of features in the rules, or specific feature in the consequent part of the rules. It can be noticed that when we increase the number of constraints, the number of association rules in the final result decreases considerably. We observe that running time also decreases as we add more constraints, but not at a significant rate.

4. Conclusion

A major task of customer relationship management department in many companies is customer churn analysis. The objective of this kind of analysis is to gain insight into consumers' behavior who are about to leave for another service company. Timely detection is believed to prevent these customers from attrition. Retaining current customers are known to take less effort and budget than acquiring new customers. The cost effectiveness is even higher if customers are valuable ones. Most business is therefore taking the customer retention issue seriously. Customer churn prediction models are a kind of tool that help marketing planners to sense the churning before it actually happens. Prediction models are conventionally built by the systematic process using statistical methods such as regression analysis. Since the emergence of new technology such as data mining, more and more business analysts have paid attention to this new technology. Many data mining methods including decision tree induction, support vector machines, rule induction, and so on, have been applied to the churn analysis task.

We propose in this paper that data mining methods suitable for business applications should not only yield high accuracy models, but they should provide comprehensible models for non-experts to understand. Black box models are obviously not easy to understand. We consider the rule induction method based on the association mining algorithm. Our proposed method of rule induction also incorporates module for users to specify constraints on the induced rules. The implementation based on the concept of constraint logic programming is illustrated in the paper. We present the design framework of our rule induction module as a part of an inductive decision support system. The completion of other supporting modules in our design is our future research direction.

Acknowledgements

This research was supported by the SUT Research and Development Fund, Suranaree University of Technology.

References

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases", Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, **(1994)**, pp. 487-499.
[2] L. Bin, S. Peiji and L. Juan, "Customer churn prediction based on the decision tree in personal handyphone system service", Proceedings of 2007 International Conference on Service Systems and Service Management, Chengdu, China, **(2007)**, pp. 1-5.
[3] C. L. Blake and C. J. Merz, Churn data set. UCI Repository of Machine Learning Databases, [http://www.ics.uci.edu/~mlearn/MLRepository.html], University of California, Irvine, **(1998)**.
[4] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction", Expert Systems with Applications, vol. 36, **(2009)**, pp. 4626-4636.
[5] S. Chaudhuri, U. Dayal and V. Narasayya, "An overview of business intelligence technology", Communication of the ACM, vol. 54, no. 8, **(2011)**, pp. 88-98.
[6] M. Darbari and N. Dhanda, "Applying constraints in model driven knowledge representation framework", International Journal of Hybrid Information Technology, vol. 3, no. 3, **(2010)**, pp. 15-22.
[7] K. W. De Bock and D. Van den Poel, "Ensembles of probability estimation trees for customer churn prediction", Proceedings of the 23rd International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, Cordoba, Spain, **(2010)**, pp. 57-66.
[8] K. W. De Bock and D. Van den Poel, "An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction", Expert Systems with Applications, vol. 38, **(2011)**, pp. 12293-12301.
[9] M. A. H. Farquad, V. Ravi and S. Bapi Raju, "Rule extraction from support vector machine using modified active learning based approach: an application to CRM", Proceedings of the 14th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Cardiff, U.K., **(2010)**, pp. 461-470.
[10] R. Fitriana, Eriyatno and T. Djatna, "Progress in business intelligence system research: a literature review", International Journal of Basics & Applied Sciences, vol. 11, no. 3, **(2011)**, pp. 96-105.
[11] A. E. Eiben, A. E. Koudijs and F. Slisser, "Genetic modeling of customer retention", Proceedings of the First European Workshop on Genetic Programming, Paris, France, **(1998)**, pp. 178-186.
[12] M. Gorgoglione and U. Panniello, "Beyond customer churn: generating personalized actions to retain customers in a retail bank by a recommender system approach", Journal of Intelligent Learning Systems and Applications, vol. 3, no. 2, **(2011)**, pp. 90-102.
[13] M. Gouider and A. Farhat, "Mining multi-level frequent itemsets under constraints", International Journal of Database Theory and Application, vol. 3, no. 4, **(2010)**, pp. 15-34.
[14] O. Isik, M. C. Jones and A. Sidorova, "Business intelligence (BI) success and the role of BI capabilities", Intelligent Systems in Accounting, Finance and Management, vol. 18, **(2011)**, pp. 161-176.
[15] W. Kowalczyk and F. Slisser, "Modelling customer retention with rough data models, Proceeding of the First European Symposium on Principles of Data Mining and Knowledge Discovery", Trondheim, Norway, pp. 4-13 **(1997)**
[16] D. T. Larose, "Discovering Knowledge in Data: An Introduction to Data Mining", John Wiley & Sons, **(2005)**.
[17] T. Mutanen, "Customer churn analysis – a case study", Research Report, No. VTT-R-01184-06, Technical Report Center of Finland, **(2006)**.
[18] S. V. Nath and R. S. Behara, "Customer churn analysis in the wireless industry: a data mining approach", Proc. of Annual Meeting of the Decision Sciences Institute, Washington, D.C., U.S.A., **(2003)**, pp. 505-510.
[19] B. Padmanabhan, A. Hevner, M. Cuenco and C. Shi, "From information to operations: service quality and customer retention", ACM Transactions on Mgt. Information Systems, vol. 2, no. 4, Article 21, **(2011)**.
[20] J. Qi, L. Zhang, Y. Liu, L. Li, Y. Zhou, Y. Shen, L. Liang and H. Li, "ADTreesLogit model for customer churn prediction", Annals of Operations Research, vol. 168, no. 1, **(2009)**, pp. 247-265.
[21] T. Sato, B. Q. Huang, Y. Huang, M.-T. Kechadi and B. Buckley, "Using PCA to predict customer churn in telecommunication dataset", Proceedings of the 6th International Conference on Advanced Data Mining and Applications, Chongqing, China, **(2010)**, pp. 326-335.

[22] M. Shahriar and S. Anam, "Towards data quality and data mining busing constraints in XML", International Journal of Database Theory and Application, vol. 2, no. 1, **(2009)**, pp. 23-30.
[23] M. Shahriar and J. Liu, "Constraint-based data transformation for integration: an information system approach", International Journal of Database Theory and Application, vol. 3, no. 1, **(2010)**, pp. 53-61.
[24] R. Srikant, Q. Vu and R. Agrawal, "Mining association rules with item constraints", Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, **(1997)**, pp. 67-73.
[25] N. B. Syam and J. D. Hess, "Acquisition versus retention: competitive customer relationship management", Working Paper, University of Houston, Houston, Texas, U.S.A., **(2006)**.
[26] W. Verbeke, D. Martens, C. Mues and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques", Expert Systems with Applications, vol. 38, **(2011)**, pp. 2354-2364.
[27] L.-S. Yang and C. Chiu, "Knowledge discovery on customer churn prediction", Proceedings of the 10th WSEAS International Conference on Applied Mathematics, Dallas, U.S.A., **(2006)**, pp. 523-528.
[28] Y. Zhao, B. Li, X. Li, W. Liu and S. Ren, "Customer churn prediction using improved one-class support vector machine", Proceedings of the 1st International Conference on Advanced Data Mining and Applications, Wuhan, China, **(2005)**, pp. 300-306.

Authors

Nittaya Kerdprasop is an associate professor at the School of Computer Engineering, Suranaree University of Technology, Thailand. She received her bachelor degree in Radiation Techniques from Mahidol University, Thailand, in 1985, master degree in Computer Science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in Computer Science from Nova Southeastern University, U.S.A, in 1999. She is a member of ACM and IEEE Computer Society. Her research of interest includes Knowledge Discovery in Databases, Artificial Intelligence, Logic Programming, and Intelligent Databases.

Phaichayon Kongchai is currently a doctoral student with the School of Computer Engineering, Suranaree University of Technology, Thailand. He received his bachelor degree in Computer Engineering from Suranaree University of Technology (SUT), Thailand, in 2010, and master degree in Computer Engineering from SUT in 2012. His current research includes Constraint Data Mining, Association Mining, Functional and Logic Programming Languages, Statistical Machine Learning.

Kittisak Kerdprasop is an associate professor and chair of the School of Computer Engineering, Suranaree University of Technology, Thailand. He received his bachelor degree in Mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in Computer Science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in Computer Science from Nova Southeastern University, U.S.A., in 1999. His current research includes Data mining, Artificial Intelligence, Functional and Logic Programming Languages, Computational Statistics.