

Predicting Churn Probability of Fixed-line Subscriber with Limited Information: A Data Mining Paradigm for Enterprise Computing

Yingying Zhang, Jiayin Qi, Huaying Shu, and Yuanquan Li
Economics and Management School, Beijing University of Posts and
Telecommunications, 10 Xi Tu Cheng Road, Haidian District, Beijing,
China, 100876

zhangying0070@gmail.com,
ssfjy@263.net,
shuhy@bupt.edu.cn,
leo8410@gmail.com,

WWW home page: <http://www.sem.bupt.cn>

Abstract. The phenomenon of subscriber churn is becoming more and more serious in the fixed-line communications industry. In order to build customer loyalty and maximize profitability in the ever-increasing competitive marketplace, a churn prediction method becomes necessary for a fixed-line services provider. However, today's researches on churn prediction in the telecommunications industry mostly concentrate on mobile services field, rarely on fixed-line services field. One prime reason is the less amount of qualified information for churn prediction in the fixed-line services providers. In response to the limitation of information, especially the incompleteness of call details and unreliability of subscribers' demographics in the investigated fixed-line services provider, we propose, design and experimentally evaluate several churn-prediction models applying three different data mining techniques (Decision tree, regression, neural network), with predictors (i.e. input variables) derived only from subscribers' contractual information and bill details. The predictors can be mainly categorized into four types: duration of service use, payment type, amount and structure of monthly service fees, change of the monthly service fees. The result shows that these limited but appropriately designed predictors can effectively predict subscribers' churn probabilities and decision tree outperforms regression and neural network in this study, with the optimal predictive and explanatory power. What's more, it also indicates that duration of service use is the most predictive predictor, and payment type and other variables of amount and structure of monthly service

Please use the following format when citing this chapter:

Zhang, Y., Qi, J., Shu, H., Li, Y., 2006, in International Federation for Information Processing, Volume 205, Research and Practical Issues of Enterprise Information Systems, eds. Tjoa, A.M., Xu, L., Chaudhry, S., (Boston:Springer), pp.589-590.

fees within different months especially the latest months are also effective predictors. According to the result that the predictors within the latest months are more effectual, we then build different decision tree models using historical data of different amounts of months. We find that with the reduction of early monthly data for prediction, the model performance index “churner captured proportion in top ranks” declines very slightly, which can be ignored. However, the amount of the data for processing and the runtime of prediction model decreases significantly. Hence, we suggest that using relatively fewer, latest months’ data to predict subscribers’ churn trends would be an effective way.

Acknowledgements

Research of this paper is funded by the National Natural Science Foundation of China (Project No.70371056) and the Key Laboratory of Information Management & Information Economy of Ministry of Education of the People’s Republic of China.

References

1. S. Hung, H. Wang, Applying Data Mining to Telecom Churn Management, The 2004 Pacific-Asia Conference on Information Systems, Shanghai, 2004.
2. C. Wei, I. Chiu □ Turning Telecommunication Call Details to Churn Prediction: A Data Mining Approach, *Expert Systems with Applications* 3(2), 103-112 (2002).
3. R. Mattersion, *Telecom Churn Management* (APDG Publishing, North Carolina, 2001).
4. W.J. Frawely, G. Piatetsky-Shapiro, C.J Matheus, Knowledge Discovery in Databases: An overview □ in □ Knowledge Discovery in Databases, edited by G. Piatetsky-Shapiro & W. J. Frawley (AAAI/MIT Press, Cambridge, 1991), pp. 1-27.
5. M.J. Shaw, C. Subramaniam, G.W. Tan, and M. Welge, Knowledge Management and Data Mining for Marketing, *Decision Support Systems* 31(1), 127-137 (2001).
6. A. Berson, S. Smith, and K. Thearling, *Building Data Mining Applications for CRM* (McGraw-Hill, New York, 2000).
7. L. Yan, R.H. Wolniewicz, R. Dodier, Predicting Customer Behavior in Telecommunications, *IEEE Intelligent Systems* 19(2), 50-58 (2004).
8. J.R. Quinlan, Simplifying Decision Trees, *International Journal of Machine Learning Studies* 27(3), 221-234 (1987).
9. C. Yi and D. Rossiter, An Input Pre-screening Technique for Control Structure Selection, *Computers and Chemical Engineering* 21(6), 563-569 (1997).
10. L. Yan, D.J. Miller, M.C. Mozer, and R. Wolniewicz, Improving Prediction of Customer Behavior in Nonstationary Environments, *IEEE Neural Networks* 3, 2258-2263 (2001).
11. W. Au, K.C.C Chan, X. Yao, A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction, *IEEE Evolutionary Computation* 7(6), 532 – 545 (2003).