KATHOLIEKE UNIVERSITEIT
LEUVEN

Mining social networks for customer churn prediction

Wouter Verbeke[1]
Karel Dejaeger[1]
Thomas Verbraken[1]
David Martens[2]
Bart Baesens[1 3]

[1] Faculty of Business and Economics, Department of Decision Sciences and Information Management, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium

[2] Faculty of Applied Economics, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium

[3] School of Management, University of Southampton, Highfield Southampton, SO17 1BJ, United Kingdom

wouter.verbeke@econ.kuleuven.be
T: +32 16 32 68 87
F: +32 16 32 66 24

www.dataminingapps.com

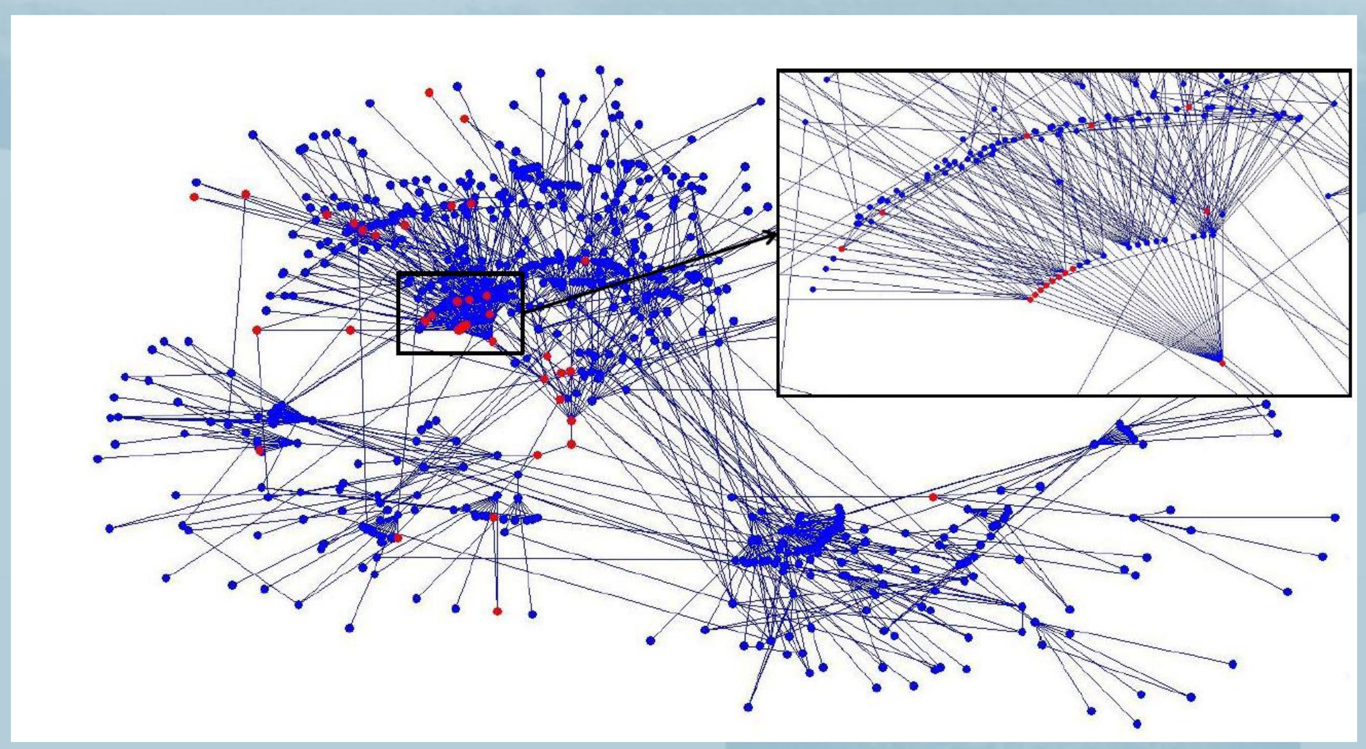
Abstract

Customer churn prediction models aim to detect customers with a high propensity to attrite. This study investigates the applicability of relational learning techniques to predict customer churn using social network information. A range of existing, extended, and novel relational classifiers and collective inference procedures have been (re-) implemented and applied on two large-scale real life data sets obtained from two international telco operators, containing both networked (call detail record data) and non-networked (usage statistics, socio-demographic, marketing related) information about millions of customers. The results of the experiments indicate the existence of a limited but relevant impact of network effects on customer churn behavior. Incorporating higher order network efffects improves the predictive power of a customer churn prediction model. Collective inference procedures however deteriorate classification performance.

1. Social network mining for customer churn prediction

Huge amounts of networked data on a broad range of network processes and information flows between interlinked entities have become available, such as for instance call logs linking telephone accounts (Dasgupta et al., 2008), money transfers connecting bank accounts, or hyperlinks relating web pages (Neville and Jensen, 2007). These massive data logs potentially hide information that is extremely valuable to companies and organizations, but as well is extremely dificult to discover due to the size and the fragmentation of the data.

Networked data present both complications and opportunities for predictive data mining. The data are patently not independent and identically distributed, which introduces bias to learning and inference procedures (Jensen and Neville, 2002; Macskassy and Provost, 2007). Relational learning aims to exploit the information contained within the network structure of data instances, and to incorporate this information within a network classification or regression model (Džeroski and Lavrac, 2001; Getoor and Taskar, 2007). The aim of this study is to apply and develop relational learners to predict customer churn using social network information derived from call detail record (CDR) data, containing a vast amount of communication logs between customers of a telecom operator.



2. Network learning systems

Macskassy and Provost (2007) introduced a node-centric, modular framework, with a network learning system consisting of:

1. a non-relational or local model,
2. a relational or network model,
3. a collective inference procedure.

This framework is adopted in this study in order to compare stand-alone versions of network learners with combinations of a local classifier and a network model, both with and without collective inference procedures. To this end, two large real life data sets have been obtained from international telco operators, with the following characteristics:

ID	#nodes	#edges	CDR time span	mean degree
1	1.365.451	2.446.672	3 months	3.58
2	8.337.285	16.494.177	6 months	3.96

Due to the large scale of the networks, a number of existing relational learners have been re-implemented using sparse and parallel computation techniques. However, the time complexity of certain relational learning techniques (such as, e.g., the network-only Bayes Classifier, (Chakrabarti et al., 1998)) appeared prohibitive for application to large scale networks (cfr. infra, future research).

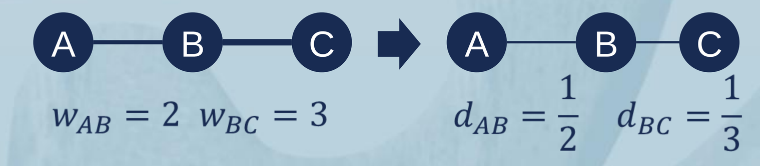
Furthermore, adapted versions of computationally expensive collective inference procedures, i.e. Gibbs sampling (Geman and Geman, 1984) and Iterative Classification (Lu and Getoor, 2003), have been developed with a reduced complexity, by making inferences concurrently for the entire network in each iteration, instead of node by node within each iteration as currently the case.

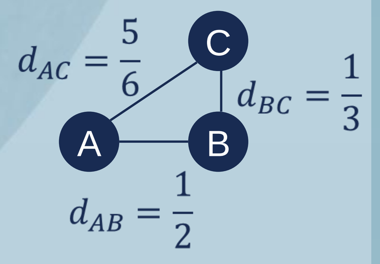
Finally, new and adjusted versions of relational classifiers have been developed to deal with the time dimension present in customer churn prediction and the skewed class distribution (~0.5% churners).

3. Incorporating non-Markovian social network effects in relational classifiers

Relational learners typically restrict the impact of the network on a node to the first order neighborhood, i.e. the nodes in the network that are directly connected to a particular node (e.g., Macskassy and Provost (2007); Neville and Jensen (2007)). However, in many applications first order Markov behavior is violated, and higher order nodes in the network have an impact which should be accounted for. Therefore, a module has been developed which can be applied in combination with any network learning technique to incorporate the impact of higher order neighborhood nodes. An adjusted version of the minimum-plus product for distance matrices is proposed, i.e. the maximum of times divided by plus product, to calculate an upgraded weight matrix that incorporates higher order neighbors as if first order neighbors.

Assume **weights** are inverse to **distances:**



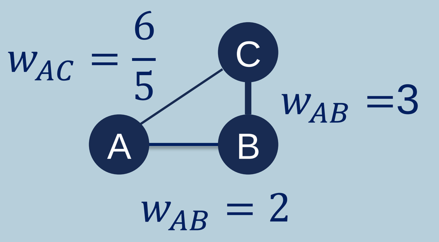


Distance to second order = sum of first order distances:

$$d_{AC} = d_{AB} + d_{BC} = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

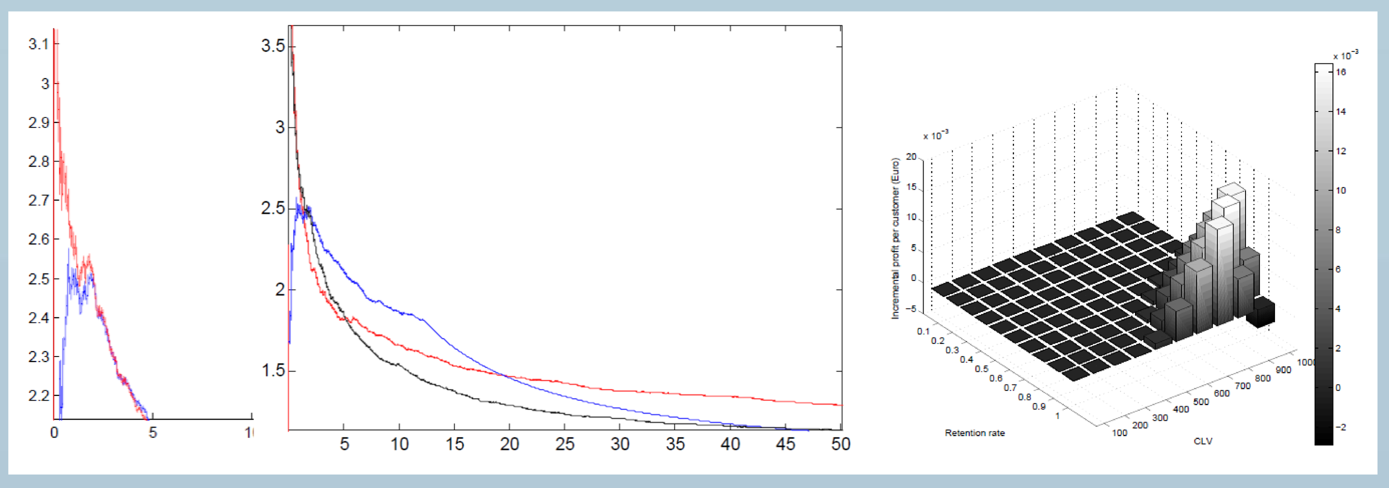
The equivalent operation for weights to summing distnaces:

$$w_{AC} = \frac{1}{d_{AC}} = \frac{1}{d_{AB} + d_{BC}} = \frac{1}{\frac{1}{w_{AB}} + \frac{1}{w_{BC}}} = \frac{w_{AB} \cdot w_{BC}}{w_{AB} + w_{BC}} = \frac{6}{5}$$



4. Experimental results

Higher order network effects improve the classification power of relational classifiers in the top fraction of customers with the highest predicted probabilities to attrite, as well in the range between 20-100%.



References

Chakrabarti, S., Dom, B., Indyk, P., 1998. Enhanced hypertext categorization using hyperlinks. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. pp. 307–319.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., Joshi, A., 2008. Social ties and their relevance to churn in mobile telecom networks. In: Proceedings of the 11th international conference on Extending Database Technology: Advances in database technology, EDBT ’08. pp. 697–711.

Džeroski, S., Lavraˇc, N., 2001. Relational Data Mining. Kluwer, Berlin, Germany.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Getoor, L., Taskar, B., 2007. Statistical Relational Learning. MIT Press, Cambridge, MA, USA.

Jensen, D., Neville, J., 2002. Linkage and autocorrelation cause feature selection bias in relational learning. In: Proceedings of the 19th International Conference on Machine Learning. pp. 259–266.

Lu, Q., Getoor, L., 2003. Link-based classification. In: Proceedings of the 20th International Conference on Machine Learning (ICML). pp. 496–503.

Macskassy, S., Provost, F., 2007. Classification in networked data. Journal of Machine Learning Research 8, 935–983.

Neville, J., Jensen, D., 2007. Relational dependency networks. Journal of Machine Learning Research 8, 653–692.