Leveraging Network Effects for Predictive Modelling in Customer Relationship Management

Christine Kiss and Martin Bichler
Department of Informatics, TU München, Germany
{kissc, bichler}@in.tum.de

Abstract: Predictive modelling and classification problems are important analytical tasks in Customer Relationship Management (CRM). CRM analysts typically do not have information about how customers interact with each other. Phone carriers are an exception, where companies accumulate huge amounts of telephone calling records providing information not only about the usage behaviour of a single customer, but also about how customers interact with each other. In this paper, we do not measure network effects, but we analyze techniques to improve classification tasks in CRM leveraging network effects. In contrast to traditional classification algorithms, we try to take into account the information about a customer's communication network neighbors in order to better predict usage behavior. The presumption in our experiment is that a customer's SMS (Short Message Service) usage also depends on the SMS usage of his social network. However, analysing huge amounts of call detail data which exhibits a graph structure poses new challenges for predictive modelling. In our work, we focus on ways to improve predictive modelling and classification leveraging data about the social network of a customer. We describe the results of an experiment using real-world data form a cell phone provider and benchmark the results against traditional approaches.

Keywords: Customer Relationship Management, Network Effects, Propositionalization, Relational Data Mining, Telecommunication Industry

1 Introduction and Motivation

Customer Relationship Management (CRM) is gaining increasing attention in many organisations. The objectives are to improve customer acquisition, customer retention, customer loyalty and customer profitability (Swift 2001; Winer 2001). Predictive modelling and classification problems are prevalent in analytical CRM. People have used logistic regressions, decision tree learner, neural networks and many other classification algorithms from the fields of multivariate statistics and machine learning for tasks such as churn prognosis, usage prognosis, and direct mail marketing. Typically, these data mining tasks involve huge amounts of data and a large number of customer attributes (Bichler et al. 2004; Rosset et al. 2001).

Nowadays, graph-based data is becoming more prevalent and interesting to the data mining community in its efforts to explore areas such as the Internet and the WWW, bioinformatics or pharmacology. Entities are modelled as graphs rather than a single table with multiple attributes. There are different approaches to mine graph data that heavily depend on the type of data available and the purpose of the analysis. CRM analysts do typically not have information about how customers interact with each other, i.e. their customer network. Phone carriers are an exception, where companies accumulate huge amounts of telephone calling records providing information not only about the usage behaviour of a single customer, but also about how customers interact with each other. Similar data might nowadays be found in the log files of Internet Service Providers, blogs or social networking sites.

There is a huge literature on network effects and social network analysis describing phenomena in networks of customers and individuals in various organizations (Hanneman 2001; Katz et al. 1994; Shy 2001; Wasserman et al. 1994). For example, the well known Metcalfe's law claims that the utility of a network equals approximately the square of the number of users of the system. One consequence of the network effect is that the purchase of a good by one individual indirectly benefits others who own the good. Telecommunication services constitute the most natural example of network externalities, since by

definition, the nature of these services involves communicating with a large number of people (Shy 2001). By purchasing a telephone a person makes other telephones more useful. This type of side-effect in a transaction is known as a positive network externality in economics (Economides 1996).

In this paper, we do not try to measure network effects, but we analyze techniques to improve classification tasks in CRM leveraging network effects. In contrast to traditional classification algorithms, we try to take into account the information about a customer's neighbors in order to better predict usage behavior. This means we analyze not only the attributes of customers but also attributes of his neighbors. The presumption is that a customer's SMS (Short Message Service) usage for instance depends on the SMS usage of his network neighbours. However, analysing huge amounts of call detail data which exhibit a graph structure poses new challenges for predictive modelling. In our work, we focus on ways to improve predictive modelling and classification based on data about the social network of a customer. We describe the results of an experiment using real-world data form a cell phone provider and benchmark the results against traditional approaches. The paper is organized as follows. In section 2 we describe approaches to analyze multi-relational data. In section 3 we present an experimental evaluation of different approaches for improving the predictive power of classification methods. Finally, in section 4 we summarize the results and describe areas of future research.

2 Classification of Multi-Relational Data

Many analytical tasks in CRM such as churn prognosis, risk management or targeted marketing involve classification of customers (e.g., being a high, medium of low risk customer). The construction of a classification procedure has also been termed pattern recognition, discrimination or supervised learning (Mitchell 1997). For example, CRM analysts of a cellular phone provider might build classification models trying to predict, whether a customer has a high, medium or low risk of switching the provider (churn management), or whether a new customer is likely to use SMS or MMS (Multimedia Message Service) a lot or not at all, based on individual attributes. Given a training and a test data set, analysts can benchmark different algorithms based on either overall accuracy (or error rate resp.), i.e. the percentage of test instances that were classified correctly, their AUC or gain curves (Bichler et al. 2004).

Traditionally, research in classification has focused mainly on attribute-value learning where each example or instance can be characterized by a fixed set of attributes for which values are given. The data set in this case can be viewed as a table where each row corresponds to an instance and each column to an attribute. The hypothesis language is propositional logic and these types of algorithms are sometimes called propositional learners (Krogel et al. 2003). While propositional learners find patterns in a given single relation, (multi-) relational data mining (MRDM) algorithms find patterns in multiple relations as they are found in today's relational databases (Dzeroski et al. 2001). MRDM is a relatively young field and as of yet, there are no standard algorithms and different domains require vastly different approaches depending on the number of relations, the volume of data and the purpose of the study.

Graphs depicting a social network can be stored in a relational data model. In our case, we have received data in a simple relational model with a table containing data about 25,478 target customers and tables containing data about their neighbours (133,042 in sum), i.e. people who interacted with the target customer either via voice, SMS or MMS. Both, ILP-based algorithms and propositionalization are potential approaches to analyze this type of data sets, and the type of method might have a considerable impact on the results.

2.1 Inductive Logic Programming and Related Concepts

Most relational data mining algorithms come from the field of inductive logic programming (ILP) (Lavrac et al. 1994), and certain derivatives. ILP systems dealing with classification tasks typically adopt the covering approach of rule induction systems (Van Laer et al. 2001). It is also possible in many ILP systems to enter known rules, which means you don't have to generate every rule but can rather define it yourself if

some common knowledge exists already. Such rules, and generally all relations except the "target" relation, are often referred to as background knowledge to be used by the analysis method (Kramer et al. 2001). Much of the art of inductive logic programming lies in the appropriate selection and formulation of background knowledge to be used by the selected ILP system. Therefore a considerable amount of expert-knowledge is necessary. In addition, ILP algorithms suffer from the high computational complexity of the task because the algorithm has to search over all the relations and all the relationships between the relations (Kietz 1993). Other MRDM techniques that are not based on logic formalisms (Getoor 2001; Knobbe et al. 1999) have been proposed, but they suffer from similar problems and are as of yet not ripe for huge data sets as is typically the case in CRM.

2.2 Propositionalization

An alternative approach is to transform multiple relations into a single one and then use traditional propositional learners to analyze the information and build a classifier. The transformation of learning problems represented relationally, into a format suitable for propositional (attribute-value) learning is called propositionalization. The goal is to generate one single relation out of multiple relations. This requires the construction of features (attributes) that capture relational properties. Logic-oriented propositionalization constructs features from relational background knowledge and structural properties. This approach performs well on structurally complex but small problems (Lavrac et al. 1991; Lavrac et al. 2002). Business databases present different challenges than those found in the classical show case areas of ILP and logic-based propositionalization, such as molecular biology or language learning. Whereas the latter often involves highly complex structural elements, perhaps requiring deep nesting and recursion, business databases are usually structurally simple (Krogel et al. 2001).

Krogel and Wrobel present an approach for feature generation by including aggregation functions, which are widely used in the database area and available in SQL to reduce the number of relations. Aggregation is an operation that replaces a set of values by a suitable single value that summarizes properties of the set. For numerical values, simple statistical descriptors such as average, maximum, minimum, and sum can be used, for nominal values the mode can be used or the number of occurrences of the different values. For feature construction domain knowledge is required to decide which attributes have to be aggregated. Examples of newly constructed features are the average number of calls a customer accepts, or the sum of neighbors of a target customer. Krogel and Wrobel experimentally showed that this approach using multi-relational aggregation outperforms a structrully-oriented ILP systems both in speed and accuracy on this class of problems (Krogel et al. 2001). However, these methods also pose problems: beside the computational cost of joins, they can produce a large number of features, and among those there are possibly higher proportions of redundant features, which might negatively impact the performance of a learning algorithm.

Unfortunately, at this point in time there is only little research investigating database-based propositionalization (Krogel et al. 2003). Dzeroski et al. (Dzeroski et al. 1999) applied a large number of relational and propositional algorithms to classification and regression variants of a learning problem and found that propositional systems performed better. This is in agreement with another evaluation by Srinivasan et.al. (Srinivasan et al. 1999) who evaluated the accuracy of propositionalization methods and ILP methods. There is also initial literature comparing database oriented propositionalization to logic oriented approaches (Krogel et al. 2003; Krogel et al. 2001). While logic oriented methods can handle complex background knowledge and provide expressive first-order models, database-oriented methods can be more efficient especially on large data sets (Krogel et al. 2003).

3 Experimental Analysis

Our data set is described by a fairly simple data structure, but a large amount of data and a large number of customers' attributes. Therefore, database-oriented propositionalization seemed particularly appropriate. Our approach follows the basic idea of RELAGGS as described in (Krogel et al. 2001). Apart from this basic procedure, there are still many degrees of freedom on how one can propositionalize multiple relations in a social network that have been the emphasis of this experiment. The goal was to predict SMS usage of customers. The attributes used for all prediction models did not include information about previous SMS usage in order to better compare different approaches.

We have derived 6 different data sets with different types of aggregate features and different ways how neighbors were selected to calculate these aggregate features and built classification models. *Model 1* was the benchmark which included only attributes of the target customer, as was the case in traditional analyses *Model 2* included also data about the voice interaction with neighbors aggregated over all neighbours of the customer. We derived a number of six aggregate features describing the interaction of a customer such as number of neighbors, average duration of calls received by the prospect, or the average number of calls received by the customer. To avoid double counting of the usage attributes of the target customer we only included interaction attributes from the neighbors to the prospect. In addition, we added information about the overall SMS usage in a three month period of these neighbours. The hypothesis is that if many neighbours of the target customer are havy sms user the target customer gets "infected". In *model 3* interaction data and the aggregate features about the usage behaviour of five closest neighbours were considered. We chose five, because the average number of degrees of a customer in the network was 5.07 (with a standard dev. of 4.7). We selected these 5 "closest" neighbours based on average communication including voice, SMS and MMS. *Model 4* is an extension of model 3 including also additional attributes of the five best contacted neighbours. Since different customers have different amount of neighbours, in *model 5* we didn't have a fixed number of five neighbours, but we chose neighbours whose communication intensity was higher than the average. The following table presents an overview of the data that was available for building the different classification models.

Neighbours / Attributes	No neighbours	All neighbours	Top 5 neighbours	All neighbours	Top 5 neighbours	Above-average neighbours
Attributes of the customer	X	X	X	X	X	X
Interaction data	X	X	X	X	X	X
Usage data of neighbors		X	X	X	X	X
Additional data of neighbors				X	X	X
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**

Table 1: Overview of the data available for model construction

We applied the C5-decision tree as propositional learner since it has performed well in previous experiments (Bichler et al. 2004; Krogel et al. 2003) and we didn't find a significant difference in the predictive accuracy compared to the multinomial logistic regression (following a Wilcoxon test with a significance level of 0.08). We used a discretized attribute for SMS usage as target attribute with values *low*, *medium* and *high*. Table 2 shows the accuracy achieved in all 6 models using a C5-decision tree algorithm.

Model 1 (Benchmark)	Model 2	Model 3	Model 4	Model 5	Model 6
63,52%	75.2%	74.01%	74.83%	73.82%	73.45%

Table 2: Results of the applied models

A statistically significant ($p>0,05$) improvement in accuracy could be reached by taking information about the network structure of a customer into account (models 2-5). The table illustrates that the best result could be reached by considering all the neighbours of the target customer (model 2 and model 4). The models with a fixed number of five neighbours do not lead to the best results, which can be explained by the fact that some customers have less than five neighbours and some have more. Between four and six (depending on the model) of the top eight attributes classified by the C5 decision tree were aggregate features derived by our propositionalization, which indicates the importance of this type of interaction data for predictive purposes. The variables with the highest influence were the SMS usage of the neighbours and the number of neighbours. Overall, the experiment illustrates the importance of network data for prediction of the usage behaviour of customers.

We also measured the impact of more than three categories of the dependent variable (SMS usage). We applied model 1, model 2 and model 3 to the same data set with the difference of using nine categories instead of three. The accuracy in all models was significantly lower: 30.5% for model 1, 43.64% for model 2 and 44.67% for model 3. But also here we could show a statistically significant improvement in accuracy by adding information about a customer's neighbourhood in the network. The data we had about customer interactions contained only information about customers of the same cellular phone provider. It should be mentioned that the goal of these experiments was not to optimize for accuracy but rather to evaluate the impact of network information on the predictive power of classification algorithms.

4 Conclusions and Future Extensions

Telecommunication services are often cited as an industry with strong network effects. In this paper we have shown, that using information about customers' neighbourhood in the call network, analysts can leverage network effects and achieve higher predictive accuracy in analytical CRM applications. The underlying hypothesis is that users with active communication behaviour also have active users in their neighbourhood. If this hypothesis is not easily transferable to other customer attributes such as churn behaviour or product preferences and needs to be analyzed in future research.

Privacy is an important issue in these types of analyses and would require a much longer discussion as is possible in this paper. In this work we have used anonymized data and focused on building classification models and not on analyzing individual interactions of customers. Therefore, we have not analyzed in detail, whether it makes a difference, whom you communicate with and what impact different individuals might have on each other.

There are a number of industries nowadays, where companies accumulate huge amounts of structured data about customer interactions. Examples are social networking sites, such as OpenBC.com, blogs, or internet service providers. Even though network effects might be less dominant in other industries, it is interesting to analyze the impact of neighbours on individual customer behaviour in other domains as well.

5 Bibliography

Bichler, M., and Kiss, C. "A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management," Tenth Americas Conference on Information Systems (AMCIS), New York, 2004.

Dzeroski, S., Blockeel, H., Kompare, B., Kramer, S., Pfahringer, B., and Van Laer, W. "Experiments in Predicting Biodegradability," Ninth International Workshop on Inductive Logic Programming, Springer Verlag, Berlin, 1999.

Dzeroski, S., and Lavrac, N. *Relational Data Mining* Springer, Berlin, 2001.

Economides, N. "The Economics of Networks," *International Journal of Industrial Organization* (14:6) 1996, pp 673-699.

Getoor, L. "Multi-relational data mining using probabilistic relational models: research summary," in: *Proceedings of the First Workshop on Multi-relational Data Mining,* A.J. Knobbe and D.M.G. Van der Wallen (eds.), 2001.

Hanneman, R.A. *Introduction to Social Network Analysis* University of California, Riverside, CA, USA, 2001.

Katz, M., and Shapiro, C. "Systems Competition and Network Effects," *Journal of Economic Perspectives* (8:Spring) 1994, pp 93-115.

Kietz, J.-U. "Some lower bounds for the Computational Complexity of Inductive Logic Programming," European Conference on Machine Learning, 1993.

Knobbe, A.J., Siebes, A., and Van der Wallen, D.M.G. "Multi-relational decision tree induction," 3rd Europ. Conference on Principles and Practice of Knowledge Disc. in Databases (PKDD), 1999.

Kramer, S., Lavrac, N., and Flach, P. "Propositionalization Approaches to Relational Data Mining," in: *Relational Data Mining,* S. Dzeroski and N. Lavrac (eds.), Springer, 2001, pp. 262-291.

Krogel, M.-A., Rawles, S., Zelezný, F., Flach, P., Lavrac, N., and Wrobel, S. "Comparative Evaluation on Approaches to Propositionalization," Springer, Berlin, 2003, pp. 142-155.

Krogel, M.-A., and Wrobel, S. "Transformation-Based Learning Using Multirelational Aggregation," Proceedings of the Eleventh International Conference on Inductive Logic Programming (ILP), Springer, 2001.

Lavrac, N., and Dzeroski, S. "Inductive Logic Programming: Techniques and Applications," Ellis Horwood, Chichester, 1994.

Lavrac, N., Dzeroski, S., and Grobelnik, M. "Learning nonrecursive definitions of relations with LINUS," Fifth European Working Session on Learning, Springer, Berlin, 1991, pp. 265-281.

Lavrac, N., Zelezný, F., and Flach, P. "Relational subgroup discovery through first-order feature constuction," Twelfth International Conference on Inductive Logic Programming (ILP), Springer, 2002.

Mitchell, T.M. *Machine Learning* WCB/Mc Graw Hill, Boston, et al., 1997.

Rosset, S., Neumann, E., Eick, U., Vatnik, N., and Idan, I. "Evaluation of Prediction Models for Marketing Campaigns," KDD 01, ACM, San Francisco, CA, USA, 2001.

Shy, O. *The economics of network industries* Cambridge University Press, 2001.

Srinivasan, A., King, R., and Bristol, D.W. "An assessment of submissions made to the Predictive Toxicology Evaluation Challenge," Proceedings of the Sixteenth International Joint Conference on Aritifical Intelligence, Morgen Kaufmann, San Franciso, CA., 1999, pp. 270-275.

Swift, R.S. *Acceleratin Customer Relationships: using CRM and Relationship Technologies* Prentice Hall, New Jersey, 2001.

Van Laer, W., and De Raedt, L. "How To Upgrade Propositional Learners to First Order Logic: A Case Study," in: *Relational Data Mining,* S. Dzeroski and N. Lavrac (eds.), Springer Verlag, Berlin, 2001.

Wasserman, S., and Glaskiewics, J. *Advances in Social Networks Analysis* Sage Publications, 1994.

Winer, R.S. "A framework for customer relationship management," *California Management Review* (43(4)) 2001, pp 89-105.