Emergent Self-Organizing Feature Maps used for Prediction and Prevention of Churn in Mobile Phone Markets

A. Ultsch

Philipps-University of Marburg

Department of Computer Science

Hans-Meerwein-Str.

35032 Marburg

In this paper we report the application of a combination of emergent self organizing Maps, U-Matrix methods and knowledge conversion (Allview) to mobile phone customer data. Aim of this approach is to discover knowledge to be used for the prediction and prevention of churn. The derived knowledge allows to construct an effective churn prediction classifier. The Allview technology produces knowledge that can be used for a clearer understanding of who the customers and in particular the churners are. Moreover possible motives for churnig can be derived. In combination with revenue data Allview is an effective tool for customer relationship management.

1 Introduction

Customer Relationship Management (CRM) means to know and understand the interaction between customers and a business [Brown 00]. CRM is a central issue in many businesses. Data Mining can be defined as the process of discovering new, understandable and useful knowledge in data sets [Ultsch 99]. Applying Data Mining to data sources created by the customer/business interaction might therefore be effectively used to acquire valuable knowledge about a customer. Of particular interest is to know who might quit being a customer and what the motives for quitting are.

Emergent self organizing maps (SOM) together with U-Matrix methods (in the following called Allview technology) allow the discovery of structures in large high-dimensional data sets [Kohonen 82]. These structures are formed by an emergent process and represent a higher level of structuring the data set. They provide a nonlinear mapping superior to classical clustering algorithms [Ultsch 95]. Chances are good that Allview shows new, formerly unknown, structures in the data set.

The structures discovered by Allview might be new, but in general they are not immediately useful for CRM. The structures by themselves do not represent knowledge. Under knowledge we understand statements about the data set that are understandable by humans, that are also interpretable by a knowledge based system and, in particular, have a meaning to the business. I.e. knowledge must lead to a non trivial understanding of important features of the data generating process. Using the knowledge conversion algorithm sig* knowledge in this sense can be extracted from the structures in emergent SOMs [Ultsch 94].

In this paper we report the application of the Allview knowledge discovery method to data from a Central European mobile phone company. Main focus was to find knowledge which can be used to predict and ultimately prevent customers to discontinue a contract with the mobile phone company.

2 Churning

Churn means the discontinuation of a contract with a business [Brown 00]. In many European countries market structures of wireless telecommunications have changed in the last years form monopolistic to very competitive markets [Edmonds 97]. To be able to control and reduce the customer churn rate may be a vital factor for the survival of mobile phone companies [Edmonds 97]. In particular it would be necessary to prevent the churning of profitable customers. If it is possible to predict whether a customer is likely to churn in two months, given today's customers record, appropriate action on the side of the business may prevent the loss of the customer. In order to prevent the churning of customers, however, it is necessary to know WHO the churners are. Furthermore it is important to know WHY a customer decides to discontinue a contract.

3 Conventional Churn Prediction

A typical approach to the problem of churn prediction is as follows: a sufficiently large data set containing churning and non churning customers is used to construct a classifier. This classifier is an algorithm that is able to decide, given a customer data set, whether the customer is likely to churn or not. These classifiers may be constructed using, for example, Artificial Neuronal Networks like Multilayer Perceptrons (Backpropagation) or Bayesian Statistics. Many Data Mining programs use decision trees constructed with heuristics like CART or C4.5 [Woods/Kyral 97].

The quality of the output of the classifier is measured in terms of sensitivity, specifity and accuracy [Griner et al 81]. The sensitivity of a classifier to predict a class c is defined as the number of data for which a correct prediction is given divided by the number of all members of class c. The specifity to predict a class c is defined as the number of data sets that are correctly classified to be not in c divided by the number of data not belonging to c. The best possible classifier has a sensitivity and specifity of 100%. In many classifiers there is a trade off between sensitivity and specifity. An increase in sensitivity will be typically accompanied by a decrease in specifity. A display of sensitivity vs. specifity is called Receiver Operating Characteristic (ROC) curve [Erkel 98]. Besides sensitivity and specifity the overall accuracy is also used to describe the quality of a prediction. Accuracy is the percentage of correct predictions. These quality measures are typically used adjust parameters of a classifier in order to find an optimal classifier [Domingos 99].

4 Problems of Conventional Churn Prediction

In a typical churn prediction approach a classifier is constructed and the measurements mentioned in the last chapter are made. Typically the first results of a classifier are not sufficient and either the parameters of the classifier are changed or another classifier model is used, This process is repeated until the quality of the prediction achieved by the classifier is sufficient. If an accuracy of round about 90% is reached for a given classifier, the ability of the constructed classifier to predict churning might be considered sufficient [Domingos 99].

A big problem is, however, to find out WHY a customer wants to churn. If it is worthwhile to keep a certain customer it is necessary to decide what business actions should be taken in order to prevent churning (CRM). For this a hypothesis is necessary on the reasons for the discontinuation of a contract. Such a piece of knowledge is, however hardly provided by the constructed classifier.

So while the constructed classifier might be useful for churn prediction it is not very helpful in churn prevention. In the following the construction of a classifier is described using knowledge extracted from Emergent SOM that overcomes these problems.

5 Mobile Phone Customer Data

The Data used for this study consisted of a sample of 300.000 customer data records. The data was provided by a Central European mobile phone company. The time span of the data was June 1999 until February 2000. We used 21 variables concerning usage of the telecommunication network. The variables described the following aspects of customer behaviour:

- money spent, for example, accounting dates for the contract,
- usage of services, like SMS,
- usage of different networks,
- destinations of long distance calls
- usage times, when the calls took place , for example daytime, night etc.

For the data an emergent SOM of dimension 128 x 128 was constructed [Ultsch 99]. With U-matrix technologies 47 groups could be found in the data. Figure 1 shows an example of such an U-

Matrix for customers that are characterised by using only domestic telephone services. As it can be seen 7 customer groups can be identified in this class.

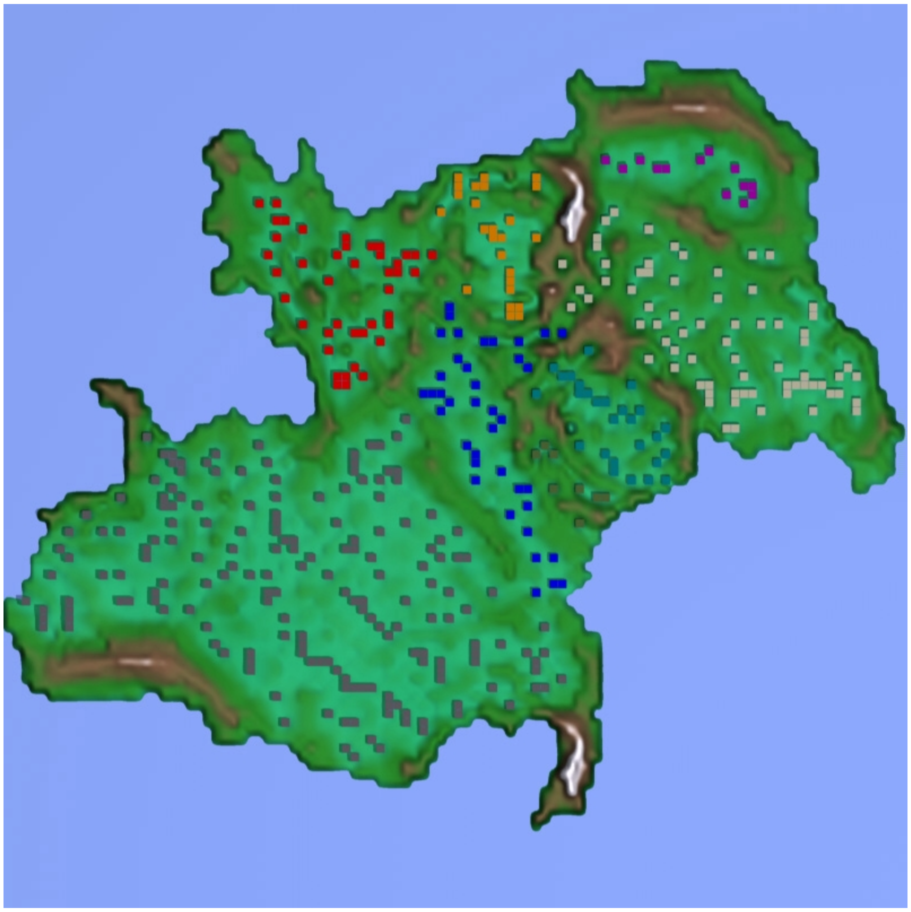

Figure 1: U-Matrix of domestic mobile phone users.

6 Knowledge Conversion

Main aim of the data mining approach discussed here is the discovery of new and useful knowledge in a data set. With the Allview technology described above groups of customers with common mobile phone usage characteristics could be identified. This by itself might be interesting but no so much as a description of what the meaning of the groups are. The algorithm sig* operates on the groups identified by Allview and produces a description of these groups in the form of understandable decision rules [Ultsch 94]. In contrast to other decision rule algorithms, like decision tree inference, the understandability of the generated is the main aim of this algorithm. With the knowledge conversion algorithm sig* rules for each of these groups could be extracted from the emergent SOMs. Sig* produces a description for each group in the form of characterising and differentiating rules. Below such a description for group 13 is given.

```
rule__4 : Class  of Customer  is_a ’13’ if for Customer holds:
          ’Var_4’    in [ 0.2, 285.5]          % (Sig=28.9191) and
          ’Var_9’    in [ 1, 223]              % (Sig=23.4303) and
          ’Var_12’   in [ 0.3, 51.4]           % (Sig=23.1201) and
          ’Var_10’   in [ 0, 4]                % (Sig=17.5463) and
          ’Var_15’   in [ 0.1, 28.5]           % (Sig=13.1106) and
          ’Var_20’   in [ 1, 22]               % (Sig=12.3796) and
          ’Var_7’    in [ 3, 11.4]             % (Sig= 9.9734) and
Customer  is_a ’13’ but_not ’23’.

rule__4__1 : Customer is_a   '13', but_not   '23'
 if for Customer holds :
          2 of  [ 'Var_3', geq 0,   'Var_11' geq 6.3, 'Var_17' geq 23.6 ].
```

A group is described using the most significant variables first. This allows to attribute a meaning to a group. For example a group could be characterised as “customers using SMS primarily at night”. Besides the description of a group in terms of variables sig* returns a measure of how significant the variable for the description of a group is [Ultsch 94]. In the example above the significance value is given at “Sig=…”. The rules for the different groups were analysed and it was found that the same variables were the most important for several groups. So these groups could be aggregated to classes with common features. The 47 groups could be aggregated to 9 classes having common usage profiles. The size of the groups is given in Figure 2.

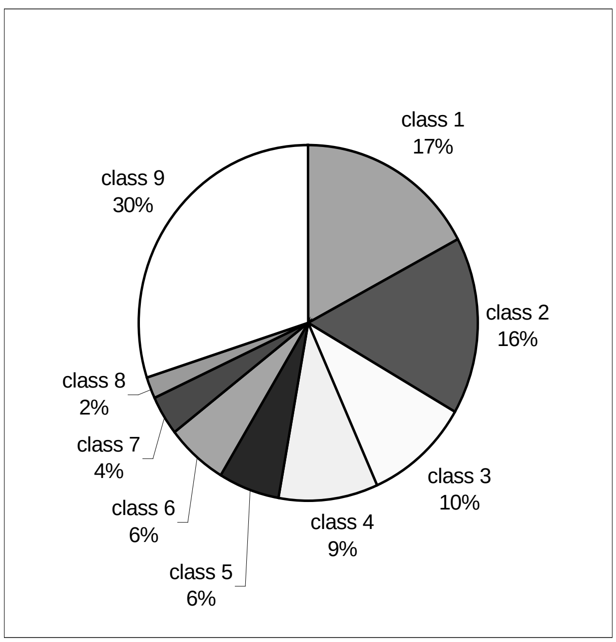

Figure 2: Classes aggregated from 47 groups

A knowledge based system can interpret the rules generated by sig*. This results in a classifier for mobile phone customers. A ROC plot of this classifier is given below. Sensitivity and specifity of the classifier is close to 100% for all rules.

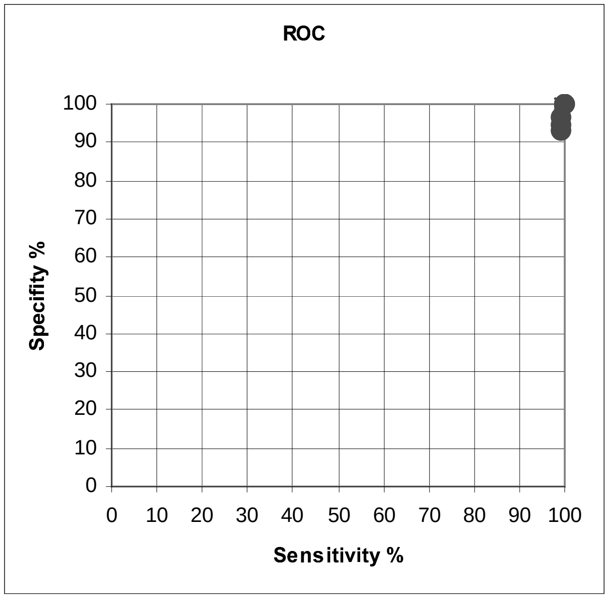

Figure 3: ROC curve of the rule based classifier for 47 groups

7 Churn Prediction

Churners were defined as customers that discontinued at least one contract with the mobile phone compny in month m. For these customers the data of the second preceding month (m-2) was selected and called Churn Data. This Churn Data was also classified using Allview.

With the methods described in the last chapters we could identify in the Churn Data the 47 groups and 9 classes detected in the data set. The rules generated by sig* for the Churn Data turned out to be effective predictors for churning. An overall accuracy for the prediction rules was measured to be 99.8%.

8 Discovered Knowledge

The description of the 47 groups respectively 9 classes in the form of rules allows to understand what type of customers are using the mobile phone network.. An attempt was made to find a description of these 47 groups using socio-economic variables. The significance values of such descriptions were to small to allow a meaningful identification of customer groups using socio-economic variables. The conclusion might be drawn that customers of the phone company are rather characterised by the usage of the network than by their social or economic background.

A comparison of the distribution of the groups among Churners and Non Churners reveals, however, interesting properties. Figure 4 shows the distribution of the 9 customer classes in the Churn Data.

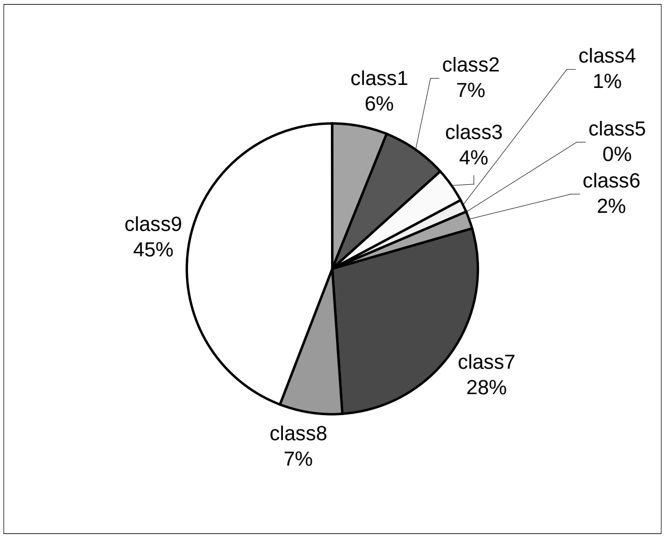

Figure 4: Distribution of classes in Churn Data

The distribution of the classes in the Non Churn Data may now be compared to the Churn Data (see figure 2). Most striking is the increase in size in class 7. The conclusion might be drawn that this type of customer is most likely to churn. Class 8 also more than doubled it size. Classes 1,2, 4 and 5 showed a significant reduction in size. Customers of these groups might be less interested in churning. Since all groups have a meaningful description in terms of network usage the reasons of a churning decision can be understood.

9 Valuable customers

It is particularly important to prevent the churning of customers who generate substantial revenues. Therefore we calculated the revenues of each of the 47 identified customer groups. Figure 5 shows the total revenue of the groups in relation to the size of the group

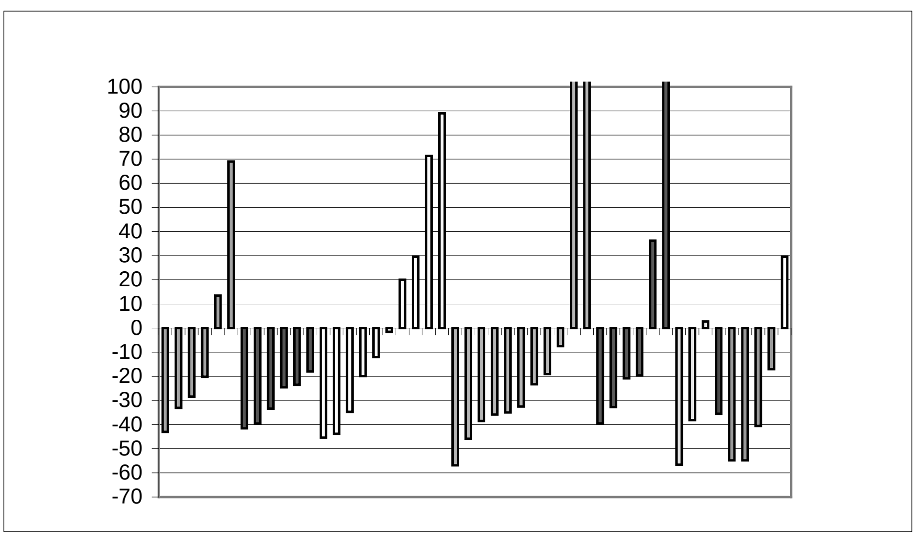

Figure 5: Revenues generated by customer groups

The zero line in figure 5 means that the members of the group contribute just as much as expected by the size of the group to the total revenue. The positive and negative numbers are percentages of additional or missing revenue in the group. Relating revenues to the groups allows a very fine grained identification of valuable customers.

It can be seen that some groups contribute substantially more revenue than it would be expected by the size of the group. Other groups miss expectations. Together with the knowledge of what the groups mean i.e. what services the customers use, this allows an efficient customer relationship management.

10 Conclusion

Self organizing maps are a non linear mapping technique that capture structures in high dimensional data not detectable by other clustering algorithms [Ultsch 95]. Emergent SOM together with U-Matrix visualisation use the phenomena of emergence to detect formerly unknown structures in data sets. In an experimental setting with real world data we could show that emergent SOM together with U-Matrix technologies (Allview) can be effectively used for the identification of groups of customers in mobile phone users. The algorithm sig* operates on the groups identified by Allview and produces a description of the groups in the form the form of understandable decision rules. In contrast to other decision rule algorithms the understandability of the generated rules is the main aim of this algorithm. The generated rules allowed to understand important characteristics of the identified groups. The rules showed to be effective predictors for churning. A comparison of the distribution of the groups for churning and non churning customers allows the inference of knowledge why customers may churn. If the groups are analysed with respect to their total revenues, customer groups can be identified for which actions to retain should be taken.

Acknowledgement

Susanne Schwenke and Ulrich Penndorf did substantial work for the project at the NeuroInformatics group at University of Marburg.

11 References

[Brown 00] Brown, S.A. (Ed) : Customer Relationship Management: A Strategic Imperative in the World of E-Business, John Wiley & Sons, 2000.

[Domingos 99] Domingos,P.: The Role of Occam's Razor in Knowledge Discovery, in Fayyad, U.(Ed), Data Mining and Knowledge Discovery, Vol 4, Kluwer Academic Publishers, pp 409-425, 1999.

[Edmonds 97] Edmonds,D.: Fair traiding in the mobile telephone market, Office of Telecommunications,http://www.oftel.gov.uk/fairtrade/mobser.htm, OFTEL, London, April 1997.

[Erkel 98] Erkel,A,R.v., Pattynama,P,M,Th.: Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology, European Journal of Radiology, 27, pp 88-94, 1998

[Griner et al 81] Griner,P.F., Mayewski,R.J., Mushlin,A.I., Greenland,P.: Selection and interpretation of diagnostic tests and procedures, Annals of Internal Medicine, 94, pp 555-600, 1981.

[Kohonen 82] Kohonen, T.: *Self-Organized Formation of Topologically Correct FeatureMaps,* Biological Cybernetics Vol. 43, pp 59 - 69, 1982

[Masand et al 99] Masand,B.M, Datta,P., Mani,P.D.R., Li,B.: CHAMP: A Prototype for Automated Cellular Churn Prediction, in Fayyad, U.(Ed), Data Mining and Knowledge Discovery, Vol 3, Kluwer Academic Publishers, pp 219-225, 1999.

[Ultsch 94] Ultsch, A.: The Integration of Neural Networks with Symbolic Knowledge Processing, in Diday et al. „New Approaches in Classification and Data Analysis", pp 445 - 454, Springer Verlag 1994

[Ultsch 95] Ultsch, A.: *Self-Organizing Neural Networks Perform Different from Statistical k-means clustering,* Gesellschaft f. Klassifikation, Basel 8th - 10th March, 1995

[Ultsch 98] Ultsch, A.: *The Integration of Connectionist Models with Knowledge-based Systems: Hybrid Systems*, Proceedings of the 11th IEEE SMC 98 International Conference on Systems, Men and Cybernetics, 11 - 14 October 1998, San Diego

[Ultsch 99] Ultsch, A.: Data Mining and Knowledge Discovery with Emergent Self-Organizing Feature Maps for Multivariate Time Series, in: Oja, E., Kaski, S. (Eds.): Kohonen Maps, pp. 33 – 46, 1999.

[Woods/Kyral 97] Woods, E., Kyral, E. *Data Mining,* Ovum Evaluates, Catalumya Spain, 1997