Data Mining in Direct Marketing - Attribute Construction and Decision Tree Induction

Petra Perner

Institute of Computer Vision and applied Computer Sciences, Leipzig

Corresponding Author: *Petra Perner, Institute of Computer Vision and applied Computer Sciences, Leipzig*

ABSTRACT

There are a lot of companies world-wide that straight collect addresses and life-style information about consumers, businesses and market places. These databases are often not set up for a special purpose and contain a lot of information. That makes data mining trickier than in the case where in the database are only collected data that are useful for making the classification model. In this paper, we have shown how decision tree induction can be used to learn the profile of an "Internet User" or "No Internet User". Attribute construction based on domain knowledge were necessary since the available database was not a special purpose database. Decision tree induction such as C4.5 and decision tree induction based on the MDL principle were used to learn the classification model. The decision-tree induction methods based on the MDL principle showed the best result in terms of error rate and explanation capability. The save of money for the mailing action can be enormous when figuring out potential customers based on data mining. Therefore, it is always better to use such kind of method instead of using all addresses for the mailing action.

Keywords: *User Profiling, Micromarketing, Data Mining, Decision Tree Induction, Attribute Construction, Mailing Action*

INTRODUCTION

Many companies worldwide straight collect addresses and life-style information about consumers, businesses and market places. The collection and maintenance of these data is not easy since masses of data have been hosted in electronic data pools with large storage capacity. The data come from life-style questionnaires consumer volunteers take part in as well as from credit card and consumer card information. In addition, webshopes or on-line campaigns' can be used to gather information about customers. Often the single customer is not deliberate that he gives out information when he legally joins a campaign and in what context this information

can be used. Despite this, the progressive customer likes to take part in these campaign's and questionnaires since he knows that these information will be used in setting up more specific products and services for his personal needs. Social media is another platform from where information can be gathered. The question is how to use all these data in order to come up with information, which make up a new quality. Often the data are very noisy and there is no a-priori label given for the final information that should be achieved based on the data.

In this paper we want to show on a specific example how these data can be used and what

tricky problems have be to circumnavigate in order to come up with information on a higher quality level. We consider here the problem of mining a model for classification in a database that has not been set up for a specific purpose. Rather than this a lot of information have been collected that can be analyzed from different direction. We use two different decision-tree induction methods for data mining. The outcome will be a structural model in form of a decision tree. One will be a binary tree and the other one will be an n-ary tree. The user can understand the model since the tree is a hierarchical rule-like form.

In Section 2, we describe related work on data mining on marketing databased. In Section 2, we explain the material used for this study. How to set up the data mining experiment is described in Section 4. Results are given in Section 5. Finally, we discuss the issue in Section 6. We conclude our work in Section 7.

RELATED WORK

Data Mining in marketing has many special needs that play no important role in other areas. A marketing database often contains a lot of information about a customer that were collected over time from a company but also data that come from other sources. These databases can be analyzed in different directions. Often it is important to know the customer profile. However, also other information should be extracted based on the company's need. The databases are characterized by the fact that they are often imbalanced, incomplete, or the class label is missing. Therefore, special methods have been developed that deal with this kind of data.

An overview about data mining in marketing is given in [1]. The data mining process is described as well as the techniques used for the different data mining tasks. Predictive models in direct marketing seek to identify individuals most likely to respond to promotional solicitations or other intervention programs. Classification techniques are used for customer identification, customer attraction, customer

retention, and customer development. Sequencing is used to relate events in time, based on a series of preceding events. Business clustering can help marketers discover distinct groups in their customer bases and characterize customer groups based on purchasing patterns.

While standard modeling approaches embody single objectives, real-world decision problems often seek multiple performance measures [2]. Decision-makers desire solutions that simultaneously optimize on multiple objectives, or obtain an acceptable tradeoff amongst objectives. Bhattacharyya [3] proposes the use of evolutionary computation based procedures for obtaining a set of no dominated models with respect to multiple stated objectives.

Prinzie and Van den Pol suggest [4] calibrate a decision model by taking into account the constraint (e.g. budget constraints) and the thresholds (e.g. only mail to customers with an expected profit higher than the investment cost) known in advance from the application. They can show that the model obtained from the data mining process is a sub-optimal model if the constraints/thresholds are taken into consideration.

The problem with imbalanced data sets in marketing is described in Lin and Lui [5]. Often only 1% of the data are positive and the rest is negative. They introduce the lift index to deal with imbalanced data sets and the claim that the predictive accuracy can-not be used as suitable evaluation criteria.

In [2] is described a product recommendation method based on customer lifetime value (CLV). The CLV of a customer is evaluated in terms of recency, frequency, monetary (RFM) variables. The analytic hierarchy process (AHP) was applied to determine the relative weights of RFM variables in evaluating customer lifetime value or loyalty. Clustering techniques were then employed to group customers according to the weighted RFM value. Finally, an association rule mining approach was implemented to provide product recommendations to each customer group. The experimental results demonstrated that the approach outperformed one

with equally weighted RFM and a typical collaborative filtering (CF) method.

An interesting study on the impact of preprocessing on data mining for marketing data presented Crone et. al [6]. This paper investigates the influence of different preprocessing techniques of attribute scaling, sampling, coding of categorical as well as coding of continuous attributes on the classifier performance of decision trees, neural networks and support vector machines. The case-based analysis provides empirical evidence that data preprocessing has a significant impact on predictive accuracy, with certain schemes proving inferior to competitive approaches.

Feeldersa et. al [7] describe the different stages in the data mining process and discuss some pitfalls and guidelines to circumvent them. Despite the predominant attention on analysis, data selection and pre-processing are the most time-consuming activities, and have a substantial influence on ultimate success. Successful data mining projects require the involvement of expertise in data mining, company data, and the subject area concerned. Despite the attractive suggestion of `fully automatic' data analysis, knowledge of the processes behind the data remains indispensable in avoiding the many pitfalls of data mining.

The problem with unknown class labels has not been studied in deep yet. Often clustering methods are applied in order to find groups of customers but that only works when the structure of the groups is already hidden in the data. That methodology cannot be applied to all kinds of questions a marketer has to analyze. We present in our paper a method how to label the data by combining several attributes in order to construct the class label from them. This method requires domain knowledge. We will show in our work how this can be done.

THE APPLICATION

Consumer databases contain a lot of information that are collected under the aspect to get as much as possible life-style information for a particular customer. There are some special

needs behind but in general, they are not collected for the construction of a particular model. In this respect, they contain more information as necessary and need to be segmented and aggregated well in order to use them for the construction of a particular model. We consider the problem of classification.

The source-sharp was a database containing more than 40000 data sets of single customers. The data set was comprised of a lot of life-style information as well as marketing information. It contained information about the products a private customer has purchased, about the kind of newspapers and journals the customer reads, income, gender, age information and many other information. For confidential reasons we are not able to describe all the data that have been collected and stored in the database. The task was to predict if a customer has a private internet access or not. This class label was not contained in the database. The database contained only information on what internet related products, what type of bank account has the customer, and what internet-related services the customer was interested.

The direct information, if a customer has an internet access or not, was not contained in the database. We like to point out that this research was going on several years ago when not so many people had personal internet access but only now, we are able to publish this research.

However, the situation we report is common to marketing aspects. From the data, which have been gathered so far, several prediction about the interest of a person in products and services have to be made. Often the request are not know before.

In this case, were the class label is not available but the customer profile has to be learnt. The data-mining expert is forced to find a way to obtain the class label by using domain knowledge and construct the class label from various attributes of the database he has so far.

It is clear that the resulting model might be heavily affected by some noise but even when it can predict a limited set of potential customers;

it is valuable to the marketing people. It gives them addresses for the mailing compain and save time and money since only to potential customers are sent out letters. The buss word is micro marketing.

In a world were a lot of different products are around and people are overloaded by advertisment material it is important to reach only those people with the advertisement that are highly interested in that offer. It can save time and costs for the mailing action. Sometimes more that 70 % of the costs compared to a mass mail action.

Table1. *Excerpt of the database with name, addresses and life-style data base*

Name	Address	Age	Sex	Children	Income	...	Journals
Mustermann	leipzig	33	Male	2	Xx	...	...
...	...	...	...	...	...	...	...

HOW TO SET UP THE DATA MINING EXPERIMENT

The Attribute Construction Problem and the Labeling Problem

In marketing experiments, there might often be no class label available. The database also might be used several times for different purposes. If a company knows what the company is interested in then the data´s are collected in such a way that the particular information are in the database. For finance institutes, health insurance institute or companies that sell special products this might be true.

Address base marketer are collecting data about customers or individuals that are available somehow independently of the selling companies particular interest in the use of the data.

That means the data can be used but the label has to be constructed by the datamining expert based on the domain knowledge he has and on the data that are available. So if we wish to label the data in internet user or not. How to do this?

The data base contained information if an individual buys computer journals, has an on-line account at a bank, bought a computer, does online-banking and so far.

In case one of the attributes matched, the data set got the label "Internet user". If none of the attributes matched the individual profile, the data set got the label "NO Internet User".

The attributes that have been taken for the construction of the label have been taken out of the database for further data mining since they correlate with the label.

For the data mining experiment was only used the label and all other information about an individual except for the attributes that have been taken to construct the label.

Application of Decision Tree Induction

The database is segmented into the new attributes and the attributes not used for the construction. The resulting subset of data have been given to decision tree induction tool Decision Master [1]. First, the standard C4.5 algorithm [2] was used and then n- ary decision tree induction methods based on the MDL principle [3] was used.

Cross-validation was used to evaluate the model. Table 2 shows the result for both methods. C4.5 perform not as good as the n-ary decision-tree induction methods.
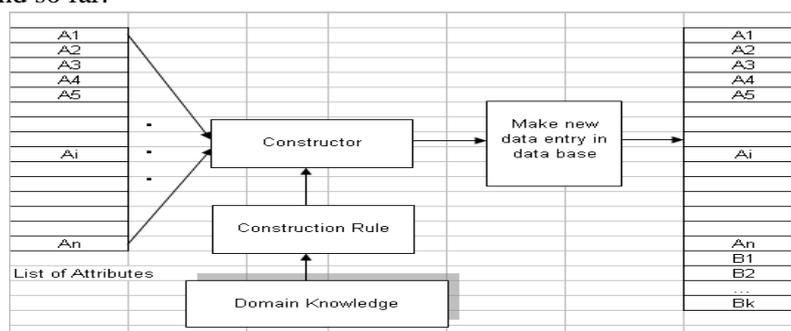
Fig1. *Construct New Attributes from Attribute List*

RESULT

The error rate of the C4.5 decision tree is much higher than the one of the decision tree based on the MDL principle (see table 2). Besides that is the explanation capability of the model of the n-ary decision tree much better to understand and made more sense to the domain expert. The n-ary decision tree was comprised of fifty-six

rules while the C4.5 decision tree was comprised of hundred four rules (see table 2). The rules of the n-ary decision tree are not so long and the tree gets not so bushy than the C4.5 decision tree.

Table2. *Accuracy and Number of Rules of the learnt Tree*

Decision tree	C4.5	DT based on MDL
Accuracy	64.45%	69.3%
Number of rules	104	56

MEASURE OF RETURN-OF-INVEST BY THE MARKETING CAMPAIGN

Based on the learnt model was done a mailing campaign. The Return of Invest was measured based on the response rate of a customer and the costs for the mailing complain.

The database was segmented based on the learnt decision tree model. The rules that described the class "Internet User" were taken to segment the database. The subset of data entries that meet these rules have been taken out and the addresses associated to the data entries were used for the mailing action. The results are shown in table 2.

Table3. *Results*

All data entries	Maling cost in Euro(0,5 Euro per Letter)	Potential internet users	Mailing cost in Euro (0,5 Euro per Letter)	Saved money in Euro
38.010	19.005	17.897	8948,50	10.056,50

DISCUSSION

Induction of a classification model based on a data set needs to have the right data for the special purpose in the database. If a special experiment is set up for the special purpose this fact is considered when preparing the experiment.

Often are collected all data that are available for one person. Such a database can be used for different purposes. However, the attributes might not describe well the right information. New attributes need to be constructed based on domain knowledge and the old attributes. How this can be done for learning a classification model based on decision tree induction that can classify customer data into "Internet User" or

"No Internet User" has been shown in this paper.

After the construction process, the database has to be segmented so that only the information, that are useful for the special purpose, and the new constructed attributes without the ones on which they are constructed are used for learning the classification model.

The most used decision tree induction methods is C4.5. The method can learn on average a good model for all kinds of data. Other methods such as n-ary decision tree induction methods can do sometimes better. We used the decision-tree induction method based on the MDL principle. The resulting tree shows a better performance based on the error rate and for the explanation capability. The rules are shorter and more compact and give a better understanding of the persons profile for "Internet User".

After the model has been learnt, the database can be segmented into the data entries of the "Internet Users". Only these people get a mailing letter. The save of money and time is enormous. Therefore, it is always better to use new techniques such as datamining methods to mine the database.

CONCLUSIONS

In this paper, we have shown how decision tree induction can be used to learn the profile of an "Internet User" or "No Internet User". Attribute construction were necessary based on domain knowledge since the available database was not a special purpose database and did not contain the class label. Decision tree induction such as C4.5 and decision tree induction based on the MDL principle were used to learn the classification model. The result of the decision-tree induction method will be a hierarchical structural model expressed by "IF…THEN…" rules. This representation allows a user to understand on what information the decision has been made. The decision-tree induction methods based on the MDL principle showed the best result in terms of error rate and explanation capability. The resulting tree is more compact than the binary tree. The rules representing the tree express much more the

domain knowledge of an expert. However, the accuracy of the model is still low. Here we observe the same as Lin and Lui [5] that the accuracy might not be good evaluation criteria. The reason for this low accuracy might be that the labels might be very noisy and that may be not all necessary features are included in the database. Nonetheless, the save of money for the mailing action can be enormous when figuring out potential customers based on data mining. Therefore, it is always better to use such kind of method instead of using all addresses for the mailing action.

ACKNOWLEDGEMENT

REFERENCES

[1] Mushtaq and H. Kanth, Data Mining For Marketing, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169, Volume: 3 Issue: 3 985 – 991

[2] Duen-Ren Liua, Ya-Yueh Shih, Integrating AHP and data mining for product recommendation based on customer lifetime value, Information & Management 42 (2005) 387–400

[3] Siddhartha Bhattacharyya, Evolutionary Algorithms in Data Mining: Multi-Objective Performance Modeling for Direct Marketing, KDD 2000, Boston, MA USA, p.465-473

[4] Prinzie, A., & Poel, D. V. (2005). Constrained Optimization Of Data-mining Problems To Improve Model Performance: A Direct-Marketing Application.Expert Systems With Applications, 630–640.

[5] Charles X. Lin and Chenhui Lui, Data Mining in Direct Marketing, KDD 88

[6] Sven F. Crone, Stefan Lessmann, Robert Stahlbock, The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing, European Journal of Operational Research 173 (2006) 781–800

[7] Feeldersa,*, H. Danielsa,b, M. Holsheimer, Methodological and practical aspects of data mining, Information & Management 37 (2000) 271-281

[8] Data Mining Tool Decision Master; www.ibai-solutions.de

[9] Quinlan, J.R. Induction of Decision Trees, Machine Learning 1, pp. 81-106, 1986

[10] Fayyad, U.M., Irani, K.B., Multi-interval discretization of continuous-valued attrib-utes for classification learning, Proc. 5th ACM SIGKDD Intern. Conference on Knowledge Discovery and Data Mining, pp. 53-62, 1999