December 2004

A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management

Martin Bichler
Technische Universität München

Christine Kiss
Technische Universität München

A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management

Martin Bichler
Internet-based Information Systems (IBIS)
Technische Universität München, Germany
martin.bichler@in.tum.de

Christine Kiss
Internet-based Information Systems (IBIS)
Technische Universität München, Germany
christine.kiss@in.tum.de

ABSTRACT

Extensive research has been performed to develop appropriate machine learning techniques for different data mining problems. However, previous work has shown that no learner is generally better than another learner. Comparing machine learning methods depends very much on the characteristics of a particular data set and the requirements of the respective business domain. Direct marketing is an important task in marketing departments, and one where machine learning techniques have been used repetedly. A systematic comparison of classifier performance can achieve considerable gains in marketing effectiveness. This case study provides an assessment of the predictive performance of different classification methods for campaign management. The evaluation of data mining methods for marketing campaigns has special requirements. Whereas, typically the overall performance is an important selection criteria, for campaign management it is more important to select the technique which performs best on the first few quantiles. This study selects candidate techniques and relevant evaluation criteria for campaign management and provides a guideline for similar comparison studies.

Keywords

Campaign management, classification, data mining, direct mail marketing

INTRODUCTION

Campaign management and in particular direct mail marketing is an area, where data mining has found many applications (Apte, Liu, Pednault and Smyth 2002). Support for managing campaigns a large number of classification techniques from the field of neural networks, statistics, and machine learning can be used to select customers who are most likely to respond to a direct marketing campaign. For example, the open source machine learning tool WEKA (Witten and Frank 2001) implements more than 60 classification algorithms. Given the wide variety of available algorithms and the volume of data modern organizations need to analyze, the selection of the right technique to use for a new problem is an important issue, and previous literature suggests that there can be considerable differences in the predictive power of the resulting models (Michie, Spiegelhalter and Taylor 1994). In addition, analyses in other domains indicate that no machine learning technique is generally better than others (Wolpert 1996). For repeated data mining tasks, it is therefore advisable to perform a structured comparison of different classification algorithms with respect to a particular data set.

Comparing different machine learning methods, however, is a non-trivial task and depends very much on the characteristics of a particular data set and the requirements of the respective business domain. In addition, comparison studies are time-consuming and one has to restrict attention to a few candidate techniques. Unfortunately, data mining literature provides little guidance in this respect. In our study, we focus on some of the particularities of campaign management. Our analysis is conducted based on a larger data set about the success of SMS (Short Message Service) campaigns amongst mobile subscribers at a large telecommunications provider. The paper introduces the candidate machine learning techniques and relevant evaluation criteria for campaign management and provides a guideline for similar comparison studies in this field.

Previous Literature

Classification occurs in a wide range of human activity. At its broadest the term could cover any context in which some decision or forecast is made on the basis of currently available information (Michie et al. 1994). The construction of a classification procedure has also been termed pattern recognition, discrimination or supervised learning (Mitchell 1997). The main historical strands of research are *statistical*, *machine learning* and *neural network* theory, which have involved different academic groups and emphasized different issues.

In the statistics community, linear discriminant analysis and logistic regression have assumed a major position as a method for classification problems. The number of methods suitable for these problems has been extended to include a range of new techniques, such as neural networks, instance-based learning, and decision tree induction. This development has led to an increase in the number of empirical comparisons of classification methods on a variety of problems. Unfortunately, the results of previous studies are often in direct contradiction, with one author claiming that decision trees are superior to neural nets or logistic regressions, and others making the opposite claim.

For example, Mingers (Mingers 1987) compared the ID3 rule induction algorithm to multiple regression. The results of this comparison favour ID3, but the comparison suffers from a number of limitations. Gilpin et al. (Gilpin and Ohlsen 1990) compared regression trees, stepwise linear discriminant analysis, logistic regression, and three cardiologists predicting the probability of one-year survival of patients who had myocardial infarctions. The methods were compared as to their sensitivity, specificity, and ROC curve. No statistical significant differences were noted between any of the methods. In another study in the medical domain by Long et al. (Long, Griffith, Selker and Agostino 1993) with over 5,000 patient records logistic regression and ID3 decision trees were compared and the logistic regression performed better. An overview and a critique of various other studies in the 1990's can be found in (Michie et al. 1994).

The European StatLog project (Michie et al. 1994) can be considered as one of the most exhaustive studies comparing around 20 methods on 20 datasets. The main conclusion was that the performance of the techniques, both in absolute and relative terms, varied considerably for different datasets. An attempt was made in StatLog to characterize the classification problem in terms of generic features of the datasets, e.g. number and types of independent variables, and number of classification categories. The aim was to classify the performance of the classification techniques on the basis of the problem types. This attempt had only limited success due to the wide variability between datasets. This situation is even more difficult because *rapid advances* are being made in all areas, machine learning, neural nets, and statistics.

So, for repeated classification tasks in organizations, as the one analyzed in this study, it is advisable to conduct a comparison study and evaluate the performance of different methods, in order to identify relationships among the predictive power of different classification methods. A number of more recent publications have started to focus on this subject (Gersten, Wirth and Arndt 2000, Perlich, 2003; Parekkat 2003; Rosset 1999; Rosset, Neumann, Eick, Vatnik and Idan 2001).

Problem Description and Data Set

Classification of dichotomous or multinomal attributes is prevalent in marketing and CRM. Churn prognosis and direct mail marketing describe two typical modelling problems, which need to be conducted repeatedly based on similar data sets. The classification problem in this study is concerned with campaign management. The goal is to identify customers who are likely to respond. The available data set summarizes the results of one previous SMS campaign at the telecommunication company 0_2 Germany. For our analysis we have randomly selected a data set of 10,054 instances, which was the basis for our comparison. The data set was divided in a training set of 5,112 instances and a test data set of 4,942 instances. The data sets include 165 independent and a dichotomous dependent variable. We've assessed thee various methods with reference to their implementation on WEKA (Witten and Frank 2001), SPSS® and Clementine® for Windows.

METHODOLOGY

As we have shown in the section 0, there are numerous different approaches and methods which can be applied to classification problems. Which one to select is a difficult question, and previous literature provides little guidance as to which methods are most appropriate for a particular dataset. For a new comparison study, the set of candidate methods is still huge. We have chosen to select a set of promising candidate algorithms, based on the results and recommendations of the StatLog project. Michie et al. (Michie et al. 1994) provide clear guidelines on which methods to include minimally in these types of comparison studies:

- One should probably always include the *logistic regression* or linear discriminant analysis, as it is sometimes best, and a standard, widely available procedure. Previous studies have found little practical difference in the performance of discriminant analysis and logistic regression. If the normality assumptions of attribute values in the linear discriminant analysis hold, it is expected to be more efficient (Harrell and Lee 1985).
- In the StatLog project, the *k-nearest neighbor method* was often the outright winner, so it would seem sensible to include kNN in any comparative studies.
- In many cases where kNN did badly, the decision-tree methods did relatively well in the StatLog project. We have decided to include the *C4.5* and *C5.0 algorithms* in our study as they are the latest generation of decision tree algorithms following ID3.

- In the StatLog project, in certain cases newer statistical procedures, such as SMART (Friedman 1984) got good results. Many of these techniques are nonparametric procedures, i.e. they can be used without assuming that the form of the underlying densities is known. Currently, SMART is not available in most commercial statistics or data mining packages.

We have decided to use the *logistic regression*, the *kNN method* and the *C4.5* and *C5.0 decision tree learner* for our study. The three techniques implement very different approaches to the classification problem, have performed well in previous studies, and are readily available in commercial statistics and data mining packages. We have not included neural networks in this initial study. The many possible neural network architectures combined with the large choice of parameter settings makes structuring neural networks a complex task. In addition, the result of neural networks is often difficult to explain to end users, which is an important aspect in our domain. The following subsections provide some brief descriptions of the selected techniques relevant to our analysis.

Logistic Regression

Logistic regression and discriminant analysis are some of the oldest classification procedures, and they are the most commonly implemented in software packages. They operate by choosing a hyperplane to separate the classes as well as possible. Fisher's linear discriminant analysis optimises a quadratic cost function whereas in logistic regression it is a conditional likelihood that is maximised. However, in practice, there is often very little difference between the two (Michie et al. 1994). Logistic discrimination is identical, in theory, to linear discrimination for normal distributions with equal covariances, and also for independent binary attributes, so the greatest differences between the two are to be expected when we are far from these two cases, for example when the attributes have very dissimilar covariances.

In our study, each complete record, corresponding to the $i-$th customer, is a vector $X_i = (X_i^1,...X_i^k;S_i)$, where X^j are some observable variables. The objective is to use the information contained in complete records to study the relationship between the vector $X = (X^1,...,X^k)$ of explanatory variables and S as the target variable, and then to predict the probability of success for incomplete records based on this relationship. No assumption is made concerning the distribution of X^j. Thus, if we do not have a good model for the distribution of X to start with, the distribution-free logistic regression, optimizing conditionally on the observed values of X_i, seems to be a logical choice. In our analysis we use the binary logistic regression implemented in SPSS®.

Decision Trees

Decision trees are based on the recursive partitioning of the sample space. Techniques for generating decision trees are also called classification or regression trees and have been developed over the past twenty years. In the machine-learning community, a number of researchers have been developing methods for inducing decision trees automatically from data sets, the best known of which are AQ11 and ID3, each of which has spawned a family of successors (e.g., C4.5) (Mitchell 1997; Quinlan 1986). In the statistics community CART is the best known approach.

The C4.5 method constructs a decision tree from a set of training objects and uses *information gain ratio* as the criterion for selecting the branching attribute. Each time, the method picks one attribute and computes the information gain ratio for this attribute. After the information gain ratio is generated for all the attributes, the data is split by choosing the attribute with highest gain ratio then continue by repeating the same procedure, choosing the attribute with highest gain ratio as compared to the remaining attributes. Multiple information-gain-ratio statistics have been used to select attributes. However, Mingers (Mingers 1987) shows that the predictive accuracy of the induced decision trees is not sensitive to the choice of this statistic. In addition, there are multiple strategies for pruning the tree once it is generated. In this study we will use the C4.5 implementation in WEKA and the commercial C5.0 implementation in Clementine®.

k-Nearest Neighbor

Instance-Based Learning (IBL) algorithms consist of simply storing the presented training examples (data). When a new instance is encountered, a set of similar, related instances is retrieved from memory and used to classify the query instance (target function). This means, in basic IBL, there is no explicit knowledge representation of a model. K-nearest neighbor, locally weighted regression, and radial basis functions are the most common IBL methods. IBL approaches can construct a different approximation of the target function for each distinct query instance that is to be classified. The kNN algorithm is the most basic of all Instance-Based Learning (IBL) methods. The algorithm assumes all instances correspond to points in the *n*-dimensional space R_n. The nearest neighbors of an instance are defined in terms of standard *Euclidean geometry* (distances between points in n-dimensional space). If most of the kNN are responders, then X_i is classified as a responder, otherwise it is classified as non-responder. Selecting the right *k* poses some difficulty and depends on the quality and amount of data, but several approaches have been suggested for this problem (Witten and Frank 2001).

COMPARISON OF METHODS IN CAMPAIGN MANAGEMENT

For the comparative trials, we followed a general procedure as was suggested in (Michie et al. 1994). In testing the accuracy of a classification rule, it is widely known that error rates tend to be biased if they are estimated from the same set of data as that used to construct the rules. Since we had a very large data set, we selected a sample of 5,112 instances as training data and a separate sample of 4,942 instances as a test data set. The predicted and true classifications on the test data give an unbiased estimate of the error rate of the classifier. The variables in our dataset have a variety of simple and complex relationships among them, providing a challenging domain for developing classification models. Before the comparison, we analyzed basic model assumptions, such as the homogeneity of covariances and multicollinearity, which are relevant to classification methods such as the logistic regression.

When dealing with marketing applications, models need to be evaluated with regard to the way they will be utilized from a business perspective (Piatetsky-Shapiro and Masand 1999; Rosset et al. 2001). When planning a campaign, one seeks to identify individuals most likely to respond to the campaign. Due to budget restrictions the number of individuals to be approached in the campaign is restricted. If in a campaign we intend to contact only 5% of the customers who are most likely to respond, it is unreasonable to evaluate a suggested model using accuracy over the full test data set. The model's performance on 95% of the population is irrelevant to the campaign goal. In addition, the cost of errors is different to other classification problems. While a false positive may only cost a stamp or a phone call, a false negative may cost losing a customer or losing a sale.

The success of a model is usually measured by the amount of responders captured within the targeted population, i.e. at certain cutoff points. A few measures capture the essence of a model's usefulness for campaign planning, from different business perspectives,

- the Response Rate (RR),
- the Lift and Gain Chart,
- the Response to Non-Response Ratio (RNR), and ROC Curve.

The first measure is the *Response Rate (RR)*, which indicates the frequency by which a responder can be expected when running a campaign:

$$RR_{(j)} = A_j / (A_j + B_j)$$

with A_j being the total number of responders and B_j being the non-responders in the *j*-th top quantile. RR is useful for calculating the expected profit from a campaign. Another measure is known as the *Lift*, which shows how much better the model prediction is relative to a random selection of the target population. The Lift measures the ratio between the $RR_{(j)}$ and the overall response rate:

$$Lift(j) = RR(j) / (A / (A + B))$$

For example, instead of reaching 5% of the responders when approaching randomly 5% of the population, a model could reach 20% of the responders by approaching the top model-scored 5% of the population. In other words, we are improving the random model by 4 times and the lift at 5% is 4. A *Gain Chart* displays the Lift in all quantiles simultaneously. The area under the curve is a measure of the models' relative ability to identify responders.

The *Response to Non-Response Ratio (RNR)* is the ratio between the percentage of all responders and the percentage of all non-responders in the top *j*-th quantile:

$$RNR_{(j)} = (A_j / A) / (B_j / B)$$

The RNR is independent of the overall response rate and allows for easy comparison among models. ROC and Gain Chart are methods for evaluating the performance of models. The ROC (Receiver Operating Characteristic) curve is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity). Another measure, which is often used to evaluate classifier performance is the *misclassification rate*, which is the percentage of entities classified incorrectly among all entities. For campaign management the misclassification rate is usually inappropriate since a campaign inherently focuses on some small sub-populations and not the entire population. For the purpose of this study, we focus on RR, RNR, Lift, and Gain Curve as the main criteria for model selection.

RESULTS

Gain Curves and Selected Statistics

For campaign management we used the Gain Curve as the primary decision aid. The Gain Curves show how much better a selection based on a particular classifier is doing compared to a random selection.

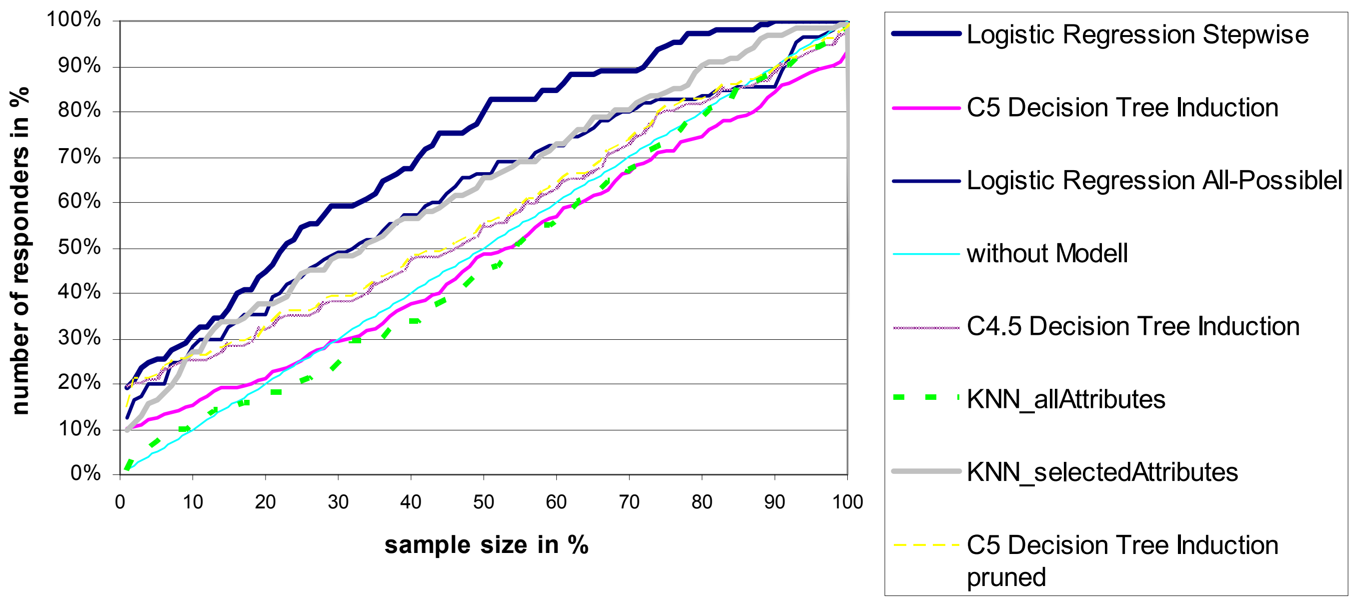

Figure 1. Cumulative Gain Chart for 6 Classification Methods

Based on the Gain Chart, it can be shown that 25.23% of the responders can be reached by addressing 5% of the total population using a stepwise logistic regression. Overall, the stepwise logistic regression did best and dominated all other methods in every percentile. The C4.5 decision tree inducer in WEKA produced good results, in particular for the first few percentiles of the Gain Chart. After fine-tuning the pruning parameters of C5.0 in Clementine, the performance mirrored the one of the C4.5 algorithm. kNN performed worst if all attributes were included. However, after including only relevant attributes (based on the stepwise logistic regression), the performance of the kNN was considerably improved.

Method Name	RR (5%)	RNR (5%)	Lift (5%)	Lift (10%)
Logistic Regression (SPSS ®; all-possible model) Most important variables based on the Wald statistic.	0,099	4,345	3,96	2,69
Stepwise Logistic Regression (SPSS®)	0,126	5,7	5,045	2,97
kNN (reduced Attributes)	0,081	3,48	3,24	2,68
C5.0 Decision Tree (Clementine®)	0,109	4,85	4,36	2,6
C4.5 Decision Tree (WEKA)	0,105	4,65	4,2	2,52

Table 1: Classifier Performance and Variable Selection

Table 2 summarizes statistics for the model selection. The response rate RR for the 5%-quantile is highest when using the stepwise logistic regression (0.126), followed by the decision trees. RNR (5%) is the ratio between the percentage of all responders and the percentage of all non-responders in the 5%-quantile. Again, stepwise regression has the best value, and kNN the worst. In terms of lift in the top 5% the stepwise logistic regression is followed by the C5.0 and C4.5 decision tree inducers. In the following, we will discuss some more detailed statistics about the individual models, their model fit and their results.

Logistic Regression

The binary logistic regression provides good results using all possible variables (*all-possible model*). Because logistic regression can not deal with non-numeric attributes these were omitted by the SPSS procedure. It is well known that linear regression models, including logistic linear regression models, become unstable when they include many predictor variables relative to the sample size. This translates into poor predictions when the model is applied to new data. *Forward stepwise selection* is an established procedure for this purpose and actually produced the best results among all classifiers. We will first describe relevant statistics for the all-possible model, and then for the stepwise logistic regression.

All-Possible Model

One possibility to analyze the significance of model coefficients of a logistic regression is the Wald statistic. It is the square of the ratio of the coefficient to its standard error. The Wald tests each coefficient, but is biased when the coefficient is large. The Omnibus tests of model coefficients provide a test of the *joint predictive ability* of all the covariates in the model. The probability of the observed results given the parameter estimates is known as the likelihood. Since the likelihood is a small number less than 1, it is customary to use -2 times the log of the likelihood. -2LL is a measure of how well the estimated model fits the likelihood. A good model is one that results in a high likelihood of the observed results. This translates to a small number for -2LL (If a model fits perfectly, the likelihood is 1, and -2 times the log likelihood is 0). -2LL is also called the Deviance (DEV). The smaller the deviance is, the better the model fits the data. The *model Chi-Square of 304.97* is calculated as the difference between the -2LL for a model with no independent variables and just a constant and the -2LL for the model with all variables. This tells us that, in terms of predicting response to SMS campaigns, the all-possible model is a significant improvement over the model with just a constant. Table 3 summarizes the results:

Sensitivity (overall percentage of responders classified as responders)	12,7%
Specificity (overall percentage of non-responders classified as non-responders)	99,4%
Predictive accuracy (p*sensitivity + (1-p)*specificity):	97,3%
Error rate (false classifications in percent of the overall number of instances)	2,69%

Table 2: Overall Statistics for All-Possible Model

Stepwise Selection

There are several procedures for variable selection implemented in statistics packages: forward selection, backward elimination, stepwise selection, and the best subset selection procedure. Stepwise selection is intuitively appealing and combines forward and backward selection. The result of stepwise logistic regression will depend on the significance level for variables entering and staying in the model and the selection criteria used. We have used the likelihood ratio test as a selection criterion for its accuracy (Menard 1995).

Sensitivity (overall percentage of responders classified as responders)	10%
Specificity (overall percentage of non-responders classified as non-resp.)	99,79%
Predictive accuracy (p*sensitivity + (1-p)*specificity):	97,6
Error rate (false classifications in percent of the overall number of instances)	2,4%

Table 3: Overall Statistics for Stepwise Regression

Decision Tree Induction

The C4.5 method performs a top-down irrevocable search and uses information gain ratio as the criterion for selecting the branching attribute. The C4.5 method in WEKA produced good results in the first 10% of the gain curve, but performed significantly worse than the stepwise logistic regression.

Sensitivity (overall percentage of responders classified as responders)	8,76%
Specificity (overall percentage of non-responders classified as non-resp.)	99,68%
Predictive accuracy (p*sensitivity + (1-p)*specificity):	97,2%
Error rate (false classifications in percent of the overall number of instances)	2,76%

Table 4: Overall Statistics for C4.5 Decision Tree

k-Nearest Neighbor Method

kNN is an instance-based learner and does not produce a model. Another disadvantage of this method is that it is computationally intensive to classify large data sets. For example, classification using our training and test data set took several hours. Nevertheless, kNN was particularly successful as a classification method in the StatLog project, and it is interesting to see how this method performs compared to logistic regression or decision tree inducers.

A crucial parameter for kNN is *k*, the number of neighbors to be used for classifying a new instance. A general rule of thumb is to use the square root of *n*, where *n* is the number of training instances (Dasarathy 1991). We found 5-NN to perform already very well for the training data set. Using all possible variables, the performance was dominated by any other method. The most important reason for this result is that the presence of irrelevant variables is always a problem with k-nearest neighbor. Second, kNN performs well with numeric attributes, but is less suitable with many nominal attributes. Therefore it is necessary to reduce the weight attached to some variables and removed others altogether, if they do not contribute usefully to the discrimination (Michie et al. 1994). There are established procedures for removing unnecessary variables in logistic regression (e.g., forward stepwise selection). In a second trial, we have only used the attributes identified as important by the stepwise logistic regression. This procedure led to considerably better results. Table 6 shows the confusion matrix for 5-NN with the reduced set of variables.

Sensitivity (overall percentage of responders classified as responders)	6,56%
Specificity (overall percentage of non-responders classified as non-responders)	99,85%
Predictive accuracy (p*sensitivity + (1-p)*specificity):	97,5%
Error rate (false classifications in percent of the overall number of instances)	2,45%

Table 5: Overall Statistics for kNN with a reduced Set of Variables

CHARACTERIZATION OF THE DATA SET

For the evaluation of our analysis, research on general guidelines for predicting preferable methods was considered. The aim of some new research is to develop pre-conditions for selecting the best model. Perlich et al. (Perlich, Provost et al. 2003) have defined a measure of the separability of signal from noise to characterize the preferable method which can be assessed by computing the area under the ROC curve.

The area under the ROC curve (AUR) is a metric for comparing classifiers across a wide range of conditions (Bradley, 1997). AUR measures the quality of an estimator´s classification performance, averaged across all possible probability thresholds; the maximum AUR can be considered as an estimation of separability of the signal from the noise (Perlich, Provost et al. 2003). Perlich et al. (Perlich, Provost et al. 2003) identified several broad patterns in the performance of tree induction and logistic regression by using data sets of very different sizes with different levels of noise. Their analysis shows that tree induction is preferable when the signal separability is high (AUR>=0.85) and logistic regression when signal separability is low (AUR<=0.8). For the calculation of the AUR we used the trapezoidal integration (Bradley 1997) to determine if the results obtained by the gain curves can be affirmed. The AUR for the C5 decision-tree is 0,76 which is less than 0,85 so the logistic regression is preferable which is consistent to our previous observations.

CONCLUSIONS AND SUMMARY

Campaign management has been described as one of the major application fields of data mining. (Apte et al. 2002). Previous work in data mining has shown that no learner is generally better than another learner. For different data sets and learing tasks people have started conducting empirical comparison studies to determine the best technique. The evaluation of data mining methods for marketing campaigns has special requirements, compared to other application domains. Whereas, typically the overall performance or error rate is the most important selection criterion, for campaign management it is more important to select the technique which performs best on the first few quantiles of the overall customer base. In addition, one typically has to deal with very large data sets. This study introduces candidate machine learning techniques and relevant evaluation criteria for campaign management and provides a guideline for similar comparison studies.

This comparison presents a number of interesting results. The stepwise logistic regression performed best and dominated all other methods (Lift(5%)=5,045). The C4.5 decision tree inducer in WEKA produced good results, in particular for the first few percentiles of the Gain Curve. K-nearest neighbor (kNN) methods provided good results only after irrelevant attributes were removed from the data set. There are a number of useful extensions one can perform in addition to the steps described in this paper. We have not considered financial parameters (e. g. savings or revenue from responders), which might be important for some applications. In addition, the breadth of methodologies used can be increased. There is a multitude of relatively new techniques, which might have benefits in certain application fields, such as various types of neural nets, meta-learning (bagging, boosting, stacking), or non-parametric statistical techniques (Friedman 1984).

REFERENCES

1. Apte, C., Liu, B., Pednault, E., and Smyth, P. (2002) Business Applications of Data Mining, CACM (45:8) 2002.
2. Dasarathy, B.V. (1991) Nearest neighbor (NN) Norms: NN pattern classification techniques, IEEE.
3. Friedman, J.H. (1984) Smart user's guide, Department of Statistics, Stanford University, Palo Alto.
4. Gersten, W., Wirth, R., and Arndt, D. (2000) Predictive Modeling in Automotive Direct Marketing: Tools, Experiences and Open Issues," KDD 2000, ACM, Boston, MA, USA, 2000.
5. Gilpin, A.E., and Ohlsen, R.A. (1990) Predicting 1-Year Outcome following Acute Myocardial Infarction: Physicians versus Computers, Computers and Biomedical Research (23) 1990, pp 46-63.
6. Harrell, F.E., and Lee, K.L. (1985) A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality, in: Biostatistics: Statistics in Biomedical, Public Health, and Environmental Sciences, P.K. Sen (ed.), North Holland, Amsterdam, 1985, pp. 333-343.
7. Long, W., Griffith, J.L., Selker, H.P., and Agostino, R.B. (1993) A Comparison of Logistic Regression to Decision-Tree Induction in a Medical Domain, Computers in Biomedical Research (26) 1993, pp 74-97.
8. Menard, S. (1995) Applied Logistic Regression Analysis Sage Publications, 1995.
9. Michie, D., Spiegelhalter, D.J., and Taylor, C.C. (1994) Machine Learning, Neural and Statistical Classification Ellis Horwood, New York, 1994.
10. Mingers, J. (1987) Rule Induction with Statistical Data - A Comparison with Multiple Regression, Journal of Operational Research Society (38) 1987, pp 347-351.
11. Mitchell, T.M. (1997) Machine Learning WCB/Mc Graw Hill, Boston, et al., 1997.
12. Parekkat, A. (2003) Operational comparison of Logistic regression, Decision trees & Neural networks in modelling mobile service churn," VIEWS Conference 2003, London, UK, 2003.
13. Piatetsky-Shapiro, G., and Masand, B. (1999) Estimating Campaign Benefits and Modeling Lift," KDD 1999, ACM, San Diego, CA, USA, 1999.
14. Quinlan, J.R. (1986) Induction of Decision Trees, Machine Learning (1) 1986, pp 81-106.
15. Rosset, S. (1999) Ranking - Methods for Flexible Evaluation and Efficient Comparison of Two-Class Models, in: Statistics, Stanford University, Palo Alto.
16. Rosset, S., Neumann, E., Eick, U., Vatnik, N., and Idan, I. (2001) Evaluation of Prediction Models for Marketing Campaigns," KDD 01, ACM, San Francisco, CA, USA, 2001.
17. Witten, I.H., and Frank, E. (2001) Data Mining Carl Hanser, München, Wien, 2001.
18. Wolpert, D. (1996) The Existence of a Priori Distinctions between Learning Algorithms, Neural Computation (8) 1996, pp 1391-1420.