What Daimler-Benz has learned as an industrial partner from the Machine Learning Project StatLog

lamreza Nakhaeizadeh
z AG, Research and Technology
360, D-89013, Ulm, Germany
9 731 505 2860, Fax: 4210
h@dbag.ulm.DaimlerBenz.COM

A

Author of this paper was co-ordinator of the Machine Learning project StatLog during 1990-1993. This project was supported nancially by the European Community. The main aim of StatLog was to evaluate di erent learning algorithms using real industrial and commercial applications. As an industrial partner and contributor, Daimler-Benz has introduced di erent applications to StatLog among them fault diagnosis, letter and digit recognition, credit-scoring and prediction of the number of registered trucks. We have learned a lot of lessons from this project which have e ected our application oriented research in the eld of Machine Learning (ML) in Daimler-Benz. We have distinguished that, especially more research is necessary to prepare the ML-algorithms to handle the real industrial and commercial applications. In this paper we describe, shortly, the Daimler-Benz applications in StatLog, we discuss shortcomings of the applied ML-algorithms and nally we outline the elds where we think further research is necessary.

1 INTRODUCTION

Before we report as an industrial partner of StatLog, on our experience gained from this project, we would like to outline the academic and practical environment from which the idea of StatLog was born in the end of the eighties in Europe.

Most of the industrial and commercial problems are either classi cation or control problems. In Statistics there are a lot of procedures which can be used to handle the practice relevant classi cation tasks. This is the same for control. A lot of classical control procedures are appropriate to be applied to solving the industrial control problems.

In the end of the eighties, some of the ML-based algorithms (symbolic ,neural and genetic) had left the research laboratories, and one could nd also a few real applications of them. Contrary to the statistical procedures, however, such ML-based algorithms and their applicability were not very well known to the industrial and commercial users. The reasons were:

- There was no algorithm-package (e.g. like the statistical packages SAS and SPSS) available to end users.
- There was no comprehensive study analysing the performance and practicability of ML-based algorithms in dealing with large-scale industrial and commercial problems.

The idea of StatLog was born in 1989 to overcome the second shortcoming, the project was supported nancially by the European Community during 1990-1993 and was co-ordinated by the author (For more information about the partners and the results see Michie et al. (1994)). The main goal of StatLog has been to compare the performance of various ML-algorithms using real applications in classi cation (supervised and unsupervised), prediction and control. At the end of the rst phase of the project, the StatLog consortium focused its activities only on the evaluation of the supervised classi cation [1] algorithms because of following reasons:

- The consortium realised that contrary to supervised learning, there is no sophisticated concept for evaluation of the quality of unsupervised algorithms. Due to this fact, unsupervised learn-

Presented at the workshop on *Applying Machine Learning in Practice* at IMLC-95.

[1] In the statistics community "classi cation" is used for both discriminance and clustering analysis which correspond to supervised and unsupervised learning. This justi es the notation "supervised classi cation". In the ML-community, however, classi cation is used to refer to learning tasks with known class values

ing was completely eliminated from the evaluation process.

The case of continuous-valued classes[2] was eliminated partly, because the number of the non statistical learning algorithms, which were appropriate for prediction, was too low. Furthermore, the prediction procedure was itself a complex task needed very quali

Dealing with con ted because almost none orithms could handle the l problem introduced to

We will refer to the al section again where we will dis ommercial applications introd Daimler-Benz. At the end of the project, the evaluation process was performed using 23 learning algorithms and 22 real applications. The evaluated algorithms, applications, evaluation methodology and the achieved results are all discussed in Michie et al (1994).

In this paper we will introduce in section two, shortly, the applications introduced to the project by Daimler-Benz. In section three, we will discuss the lessons we have learned as an industrial partner from StatLog, and we will outline how the achieved results have effected our application oriented research in the eld of ML since 1993 in Daimler-Benz.

2 DAIMLER-BENZ APPLICATIONS IN STATLOG

Daimler-Benz is an international concern with various activities not only in its traditional area car production but in di erent other elds like satellite and aircraft production, defence technology, microelectronics, information technology and di erent services like insurance, leasing and marketing. Due to this fact, the potential applications of ML-technology in Daimler-Benz is huge. In following, we will describe, shortly, the applications which were introduced to the project StatLog by Daimler-Benz.

2.1 FAULT DIAGNOSIS OF AUTOMATIC TRANSMISSIONS

The rst application deals with fault diagnosis of automatic transmissions and was realised within the project StatLog for Mercedes-Benz. The functional model of an automatic transmission is too complicated and it is not possible to use a model-based approach for fault diagnosis. Symptom-based diagnosis is here a better approach. As we started to work on this application the fault diagnosis was done in Mercedes-Benz by using a classical rule based expert system. It was developed by spending about 60 men months and by manual analysing of about 30,000 printed protocols. These protocols have led to 3000 production rules.[3]The aim of the application that we have done within the project StatLog was to show the practicability of the symbolic ML-algorithms to the end users and to convince them that ML would be a cheaper and faster approach to develop the next generation of the fault diagnosis systems in their department. Regarding the fact that daily about 1000 automatic transmissions were examined by Mercedes-Benz, availability of data was guaranteed and an inductive symptom-based diagnosis was possible. The results are reported in Michie et al. (1994), in detail.

2.2 LETTER AND DIGIT RECOGNITION

A part of Daimler-Benz, the AEG company produces address reader machines. Traditionally, they apply classi cation tools for letter and digit recognition based on non-linear regression. They were wondering whether ML-algorithms can improve the performance. The interesting aspect of their applications was the huge amount of data, one of them consists of about 58,000 examples and 50 attributes. It was for the project StatLog of interest to examine whether the available ML-algorithms can handle this huge amount of data at all .

2.3 CREDIT-SCORING

Credit-Scoring nds di erent applications in Daimler-Benz concern. Besides the Mercedes-Benz Lease-Finance interested in estimating of the risk degree of its leasing contracts, credit-scoring is in interest of debitel Kommunikation, another part of Daimler-Benz. This company o ers mobile communication services and has at present more than 300,000 customers expanding rapidly.

The traditional approach of granting credit is the so called "Judgmental" approach. According to his past experience and using a conceptual framework based on some characteristics like income, occupation and residential stability, a credit o cer has to decide, subjectively, about granting or reject of a credit. Such judgmental methods su er, however, from drawbacks as:

[2] According to one of the reviewers the correct notation for "continuous-valued classes" is "learning numeric functions". Perhaps "learning continuous-valued functions" would be a more precise notation because a "numeric function" is not necessarily a "continuous valued function" e. g. prediction the number of children is learning a "numeric function" but not "learning a continuous-valued function".

[3] One of the reviewers has suggested that having this rule set, some knowledge intensive approaches (e.g. ILP-algorithms) would be more appropriate than knowledge-poor algorithms we have used. That may be so. But as mentioned above, generating these rules is a very expensive task, generally.

High costs of training and employing of credit officers

Often wrong decision made by credit officers

For the same cas[illegible]rs have often different jud[illegible]

These difficulties justif[illegible]ic credit management systems. [illegible] we had four credit-scoring app[illegible]ance of the ML-algorithms we[illegible]his fact, in the debitel-project w[illegible] approach. We will report in section three, the shortcomings of the available ML-algorithms, especially, in dealing with the costs of misclassification and dynamics aspects of payment behaviour of the customers.

2.4 PREDICTION OF THE NUMBER OF REGISTERED TRUCKS

Prediction of development of time series is a very important basis for making management decision models. To support the management in developing the short and long term production plans, the Marketing Department of Mercedes-Benz AG in Stuttgart predicts the annual development of number of registered cars and trucks for more than 80 countries. They are heavily trying to improve the accuracy of their total market forecasts and are interested in short (quarterly, annual) and long term (five to ten years period) prediction.

In the project StatLog, our group has investigated together with the Marketing Department of Mercedes-Benz the possibilities of the application of ML-algorithms to predicting of development of trucks market. After some initial evaluations, it turned out that the predicting performance of some ML-algorithms are comparable with the sophisticated statistical approaches like regression technique. Furthermore, regarding the fact that the output of the ML-based prediction system is in the form of "if then" rules, it is very easier to involve the background knowledge in forecasting process. Regarding these advantages, within a joint project, a three-component forecasting system based on machine learning and statistical approaches has been developed. We have reported about this project in the workshop "Fielded applications of ML" organised 1993 by Pat Langley and Yves Kodratoff at the 10. IMLC in Amherst (See also Nakhaeizadeh and Reuter (1994)).

3 IMPACT OF THE RESULTS OF STATLOG UPON THE ML-RESEARCH IN DAIMLER-BENZ

The above described applications introduced by Daimler-Benz, and the other applications introduced by the other industrial partners of StatLog were together very instructive for the group of Machine Learning of Daimler-Benz. The lessons we have learned by our participation in StatLog have influenced our application oriented research since 1993. In following we will discuss some of the important aspects

3.1 LEARNING FROM LARGE-SCALE DATASETS

In the project StatLog, we had a lot of large-scale applications with high number of examples and attributes. Letter and digit recognition discussed in section 2 belong to such cases. On the other hand, in practice one encounters datasets which are even more larger than the StatLog datasets. For example, Mercedes-Benz has a dataset includes technical and commercial information about seven millions cars and trucks. In the project StatLog, we have realised that a lot of learning algorithms can not handle such large-scale datasets, and that the ML-Community needs more research in this field. Due to this fact, we work since Jan. 1995 on another rather large project called "Data Mining by means of Machine Learning". In following we will refer to it as DM-Project. A part of our research in this project is devoted to developing of ML-algorithms with ability of dealing with very large-scale databases. On the other hand, we want to examine the applicability of some pre-processing approaches to reducing the volume of data before they are used by the learning algorithms. Results from the statistical sampling theory and from the procedures are suggested by the ML-Community (e.g. windowing by Quinlan (1986) or IBL by Aha et al. (1991)) can be used as basis for our further research. The similarity concept which is one of the main pile of CBR can help as well. Especially, we think that "data prototyping" is a promising approach to reduce the volume of the data. More research in constructive induction is necessary as well to reduce the dimension of the attribute space.

3.2 NEW EVALUATION MEASURES

In supervised learning, accuracy (error) rate is a very usual criterion for evaluation of different learning algorithms. In the project StatLog we have distinguished that this measure is not always an appropriate one. For example, the applications we had in the field of credit-scoring (see section 2) have shown us that the cost of misclassification is a more important measure.

Classification a bad risk as good would cost, generally, more than classification a good risk as bad. Most of the ML-algorithms we have used in StatLog could not handle the cost matrices at all. We have realised in StatLog that more research is necessary in this field. Due to this reason, our group has done some work focusing on the cost aspects of misclassification in pruning of decision trees (see Kno[illegible] (1994)). On the other hand, we have seen in [illegible] the algorithms with very good [illegible] applicable to real industrial and [illegible] because of high computing tim[illegible] For example the statistical al[illegible]ghbours was the best algorithm [illegible] applications. But it was not [illegible] because of the high computing [illegible]oblem is meanwhile known to t[illegible] community, (see e.g. Dasarathy (1991)) we think that more research is necessary. This topic is very related to the concept "data-prototyping" discussed in section 3.1

Another problem we encountered in the project StatLog is the evaluation of unsupervised learning algorithms. As it is mentioned before, we have distinguished that contrary to the case of supervised learning, there is no sophisticated criterion for evaluation of unsupervised learning algorithms. Some suggestions are made in literature both in the statistical and ML communities e.g. Fisher (1989). Some works are going on. Developing of more sophisticated evaluation measures is one of the topics in the project DM-Project in the next two years.

3.3 ML PREDICTION PROCEDURES

As it is mentioned already, we have not performed a comprehensive evaluation of the different prediction algorithms in StatLog. The main reason has been the fact that the number of ML-algorithms which can handle the continuous-valued functions was too low. Besides CART, only NEWID could learn the continuous-valued functions. Recently, the algorithm M5 is implemented by Quinlan (1992) which can perform this task as well. Of course, the task of learning continuous-valued functions has been studies thoroughly in the statistics literature. The classical regression analysis and the K-nearest-Neighbours are typical examples. Some of that work has seeped into the ML-Literature e.g. Atkeson (1989) and Kibler et al. (1989). Another possibility to handle this task would be to use the scientific discovery algorithms. This algorithms focus, however, often on the cases with one or two attributes and are not appropriate at all to handle the large-scale applications we had in the project StatLog (e.g. one application with 256 attributes). Between the decision tree-based algorithms applicable to learning continuous-valued functions, however, only NEWID and CART were known to StatLog Consortium at the beginning of the project. All in all, we think that more research in this field is necessary. Multistrategy approaches (e.g. Quinlan (1993), Nakhaeizadeh and Westphal (1995) and Henery (1995)) might be very useful.

3.4 ML CONTROL ALGORITHMS

At the beginning of the project StatLog, as we started to analyse the available ML-algorithms, we realised that most of them can deal only with supervised and unsupervised aspects of learning but not with the control aspects. An associated partner of Daimler-Benz has introduced to StatLog a control problem with about 45 parameters concerning the position control of a communication satellite. We have determined that the available learning algorithms partly could not handle this problem and partly could deliver results worse than the results of simple classical control algorithms. Meanwhile, one can find in literature, especially in neural networks, contributions to control problems by using reinforcement learning. The applications used in such works are, however, often limited to toy examples like pole balancing. Attempts to handle practice relevant control problems can be a very interesting topic for the ML-community.

3.5 DYNAMIC ASPECTS OF LEARNING

In the classical supervised learning, the available examples are usually used to learn a concept which can be applied later to determine the class of new unseen examples. In many practical situations in which the environment changes, this procedure works no more. In the project StatLog, we encountered often this situation. In the case of fault diagnosis of automatic transmissions described in section two, the generated rules from a certain learning dataset are no longer valid e.g. after six months, due to the changes of the parameters of the diagnosis device. It has been the same in the case of credit-scoring application as well. On one hand, the payment behaviour of the customers remains no more constant over the time. On the other hand, the credit institutes change their credit policy due to changes of the credit market.

Generally speaking, application of a learned concept to classify new unseen examples will be problematic if one of the following events occur in the "out of learning periods":

- Number of attributes changes. We have experienced in the case of the credit-scoring that certain information of customers can be no more stored due to changes of legal regularities. On the other hand, it has been also the case that one is able to get more information as before about the customers e. i. one can consider more attributes.

- Number of attributes remains the same but the interpretation of the records of the datasets changes over the time. As it is mentioned above, the changes of the payment behaviour of customers in the case of credit-scoring or the changes of the diagnosis device in the case of automatic transmissions are examples for this problem.

Put together, we have learned from StatLog that the extracted rules or depe[illegible] re valid over the time due to di[illegible] n occur. Following solutions ma[illegible] roblem.

Relearning. This s[illegible] l shortcomings. On one [illegible] sible to use the informati[illegible] dataset. On the other hand [illegible] time to get again enough [illegible] vn class values for relearning. For example, in the case of the credit-scoring, it will take months before one can distinguish whether a new customer is a good or a bad risk.

Incremental learning. There are some attempts in literature to apply incremental learning to capture the dynamic aspects of the datasets (see e.g. Utgo (1989)). These contributions, however, can not cover the above mentioned aspects.

Capturing of the dynamic aspects of the datasets is not only an speci cs problem of symbolic ML-algorithms. We think that further research is necessary as well to make able the statistical and neural learning algorithms to handle such dynamic aspects. This is the subject of one work package in the Daimler-Benz research project DM-Project.

4 CONCLUSIONS

Project StatLog has shown the main shortcomings of the ML-algorithms in dealing with the real industrial and commercial large-scale applications. As an industrial partner and contributor we have learned a lot of lessons from this project which have a signi cant in uence on orientation of our research since 1993 in Daimler-Benz. We have distinguished that a lot of ML-algorithms are very good in dealing with supervised learning tasks and can be regarded as strong alternatives to neural nets and classical statistical procedures. But in many cases, they are not able to ful l the industrial and commercial needs. Especially, we have realised hat we need more research to prepare the ML-algorithms so that they will be able to handle: large-scale datasets, prediction as well as control problems and dynamic aspects of datasets. Moreover, we have seen that additional research is necessary to establish new evaluation measures. The project DM-Project which is running since Jan. 1995 in our group aims to contribute to some of these aspects.

Acknowledgement

The author would like to thank the reviewers for their useful comments ,especially concerning the section 3.

References

Aha, D. W., Kibler, D., and Albert, M. K. (1991). Instance-based learning algorithms, Machine Learning 6, 1, 37-66.

Atkeson, C. (1989). Using local models to control movement. in Touretzky . D. S. (Ed.), Advances in Neural Information Processing Systems. San Mateo, CA, Morgan Kaufmann.

Dasarathy, B. V. (Ed.) (1991). Nearest Neighbour(NN) Norms. NN Pattern Classi cation Techniques. IEEE Computer Society Press.

Fisher, D. H. (1989). Knowledge Acquisition via incremental conceptual clustering. Machine Learning, 2, 139-162.

Henery, R. J. (1995). Combining Classical Procedures. Technical Report, University of Strathclyde.

Kibler, D. Aha, D. W. & Albert, M. (1989). Instance-based prediction of real-valued attributes. Computational Intelligence, 5, 51-57

Knoll, U. Nakhaeizadeh, G. and Tausend, B. (1994) Cost-sensitive Pruning Methods for Decision Trees. In: Brazdil, P.(Ed.). Proceedings-Band of European Conference on Machine Learning (EMLC-94). Springer Verlag.

Michie, D. Spiegelhalter, D.J. and Taylor, C.C. (Eds.) (1994). Machine Learning, Neural and Statistical Classi cation. Ellis Horwood, London.

Nakhaeizadeh, G. and Reuter, A. (1994). Application of Machine Learning to Predicting Activities in the Automobile Market. Technical Report, Daimler-Benz, Ulm.

Nakhaeizadeh, G. and Westphal, M. (1995). Combining of statistical and other learning methods and its application to prediction of German Stock Price Index. To appear.

Quinlan, R. (1986). Induction of decision trees. Machine Learning, 81-106.

Quinlan, R. (1992). Learning with continuous classes, Proceedings 5th Australian Joint Conference on Arti cial Intelligence, Singapore, World Scienti c, 343-348.

Quinlan, J. R. (1993). Combining instance-based and model-based learning. In: Proceedings of the Tenth International Conference on Machine Learning, 236-243. Morgan Kaufmann Publishers.

Utgo , P. E. (1989). Incremental induction of decision trees. Machine Learning, 4, 161-186.