Forecasting of Options in the Car Industry Using a Multistrategy Approach

Stefan Ohl

Daimler-Benz AG, Research and Technology, F3S/E
P.O. Box 2360
D-89013 Ulm, Germany
ohl@dbag.ulm.DaimlerBenz.com

Abstract

In recent years, car options like air-conditioning, automatic gears, car stereo, power windows, sunroof etc. were getting more and more important for car manufacturers. Especially at those car manufacturers which offer cars individually according to customer requirements, the options affect 30% - 40% of the parts by now. Therefore, very detailed planning and forecasting of options becomes more and more important. Because customer behavior concerning the options varies from model to model and from country to country, it seems necessary and reasonable to forecast each option for each model and each country separately. The resulting huge number of data sets requires an automatic forecasting tool that adapts itself to the actual data sets and that requires almost no user interaction. Because depending on the characteristics of a time series the quality of the forecasting results varies a lot, and because "N heads are better than one," the basic idea is to select in a first step the most appropriate forecasting procedures. This selection is done by a decision tree which is generated by using a symbolic machine learning algorithm. Those selected forecasting methods produce different results that are in a second step combined to get a common forecast. In this approach there are integrated univariate time series used in the first step for running the prediction as well as symbolic machine learning algorithms for generating the decision trees as well as multivariate statistical methods and neural networks used in the combination step.

Introduction

Relevance of Options and Extras

In recent years, options of a car like air-conditioning, automatic gears, car stereo, power windows, sunroof etc. have become more and more important for car manufacturers (Dichtl et al., 1983). On the one hand, big business around the options and extras arises, but on the other hand, the huge number of extras offered optionally causes high costs as well. Mostly, these enormous costs result of a diversity of different parts and production processes that are influenced by the different options (Brändli et al., 1994; Brändli and Dumschat, 1995). In particular, at car manufacturers that offer cars individually according to customer requirements, the options affect 30% - 40% of the parts by now. Therefore, very detailed planning and forecasting of options becomes more and more important.

Situation at Mercedes-Benz

In the case investigated in this paper, the car manufacturer Mercedes-Benz offers about 400 different options for more than 100 different models structured into 4 different classes (C-, E-, S-, and SL-class) sold in more than 250 countries. In average, each customer orders about 10 - 20 options per car. Based on almost 600,000 produced cars per year, in average, only 1.4 cars are identical (Hirzel, 1995). That means there exist only a few identical cars, and in most cases these cars belong to bulk buyers like rental car companies.

Recently, even an increase in variety of models and options is expected, since new classes never offered before by Mercedes-Benz will be presented (Stegemann, 1995). Moreover, already existing classes will be extended by additional models, and completely new options like automatic windshield wipers or different kinds of navigation systems will be sold.

Analysis of Customer Behavior: Cumulated vs. Single Data

Ad Hoc Analysis

Since some options are not available for each class and because identical options differ quite often from class to class (e.g. air conditioning of C-class is different to air conditioning of E-class), it is necessary to differentiate extras per each single class. With regard to monthly

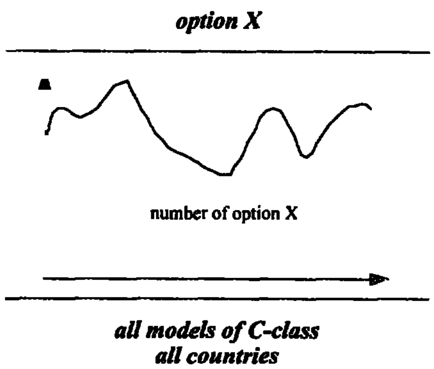

Fig. 1: Time series of single options.

cumulated data for about two years (e.g. all automatic gears sold per month in the C-class), cumulated time series of single options that represent absolute selling figures do not contain any systematic pattern as shown in figure 1. Therefore, it seems to be very difficult to forecast such kinds of time series as monthly data to a prediction horizon of 12 or even more months.

Utilizing Descriptive Statistics

However, looking at the data in a more detailed way using descriptive statistics some systematic pattern may be discovered. By analyzing the distribution of single options over different models and different countries where these options are sold, a lot of differences concerning the installation rate of an option to the total number of cars sold occur. For the options air conditioning (AC), automatic gears (AG), power windows (PW), heated seats (HS), sunroof (SR), and traction system (TS) this analysis for Mercedes C-class (Germany), and for the model C 220 (world wide), respectively, is presented in figure 2.

As it is shown in the left part of figure 2, for example, if more models C 280 and less models C 180 are produced in comparison to the previous month, the cumulative absolute number of air conditioning will increase, whereas the installation rate of air conditioning for each model will not fluctuate unsystematically a lot. Similarly, if in the current month more cars of the type C 220 are produced for Switzerland (CH) and less for Sweden (S) in comparison to the previous month, the cumulative absolute number of traction systems will increase and the cumulative installation rate of heated seats will decrease, whereas the single installation rate per model and country for these options will not vary randomly a lot (see right part of figure 2).

Based on this analysis, we detected that customer behavior concerning options varies a lot from model to model and from country to country (Ohl, 1995). Therefore, it seems reasonable to forecast and plan each single option as an installation rate for each model and each country where the option is available separately.

The less systematic kind of time series mentioned in the previous subsection of ad hoc analysis and presented in figure 1 is mainly due to the monthly shifting cars from one country to another country or from one model to another model as described before. Whereas the total number of cars produced per month is more or less stable, the number of one single model produced for one single country may vary much more (Hirzel and Ohl, 1995).

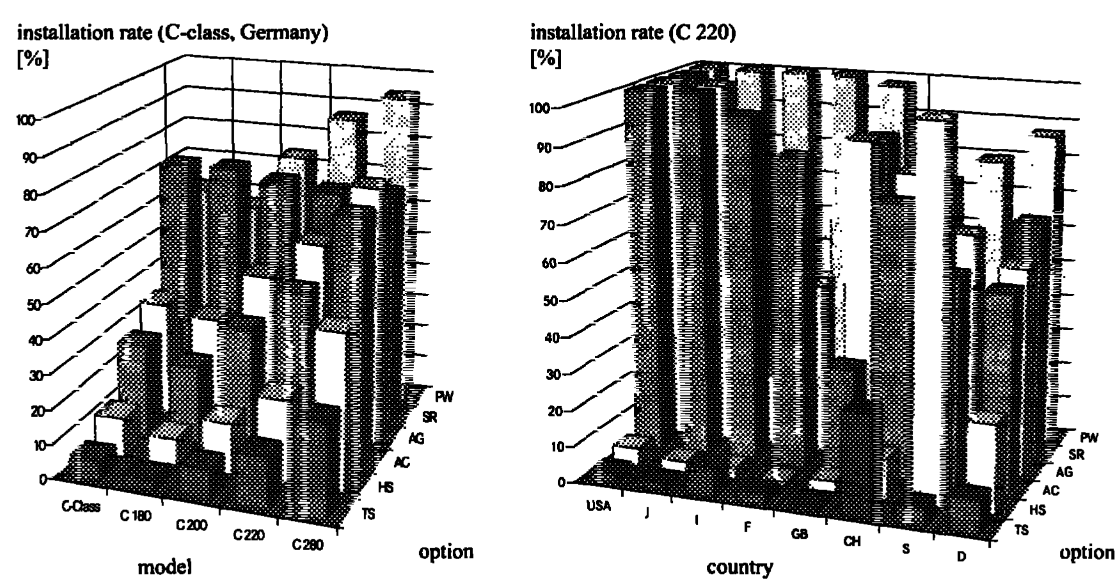

Fig. 2: Installation rates of different options at different models (Mercedes C-Class, Germany) and different countries (Mercedes C 220, world wide).

But this variation from one month to another month concerning one country or one model is in most cases planned and therefore well known, because there are seasonal effects, for instance, varying from one market to another market or from one model to another model. During summer time for instance, the number of convertibles sold is much larger than during winter time, and in markets like North America the selling figures increase when the first cars of a new model year are presented and sold in fall.

In contrast to the cumulative time series regarded in the previous subsection, analyzing single time series that represent not the absolute number but the installation rate of an option per model and country shows much more regular and systematic behavior as illustrated in figure 3.

Obviously, forecasting the single options as installation rates per model and country and multiplying these installation rates with the more or less well known planned absolute numbers of cars to be sold per model and country will yield better results than forecasting the absolute selling figures of options. This idea is illustrated in figure 3 where the time series of figure 1 is shown as the result of forecasting the single installation rates per model and country multiplied by the planned absolute number of cars as described before.

A Multistrategy Forecasting Approach

The resulting huge number of data sets requires an automatic forecasting and planning tool that adapts itself to the current data sets and that requires almost no user interaction. Another need concerning such a tool is performance, i.e. algorithms should not be too complex to avoid long computation times for doing the forecasts. The reason is that the forecasted selling figures of options are needed in a fixed company wide planning time table for further steps of the total planning process such as planning capacities of the different plants or planning the disposition of materials and parts for the next months (Ohl, 1996). This is due to the fact that the time needed for replacement of some parts and materials is longer than delivery time customers would accept. Therefore, frequently in the automotive industry it is necessary to do the disposition of parts and materials that will be delivered by internal and external suppliers before customers place their real orders.

Selection of Existing Methods

Classical Statistical Approaches. In general, four different ways of quantitative forecasting are yet well established (Makridakis et al., 1984):

- Purely judgmental or intuitive approaches.
- Causal or explanatory methods such as regression or econometric models (multivariate approaches).
- Time series methods (extrapolative univariate).
- Combinations of above mentioned techniques.

Purely judgmental or intuitive approaches are not appropriate because the user has to forecast each time series manually. The multivariate approach does not work at the application presented in this paper as well, since it is almost impossible to identify impact factors for different options for different countries. For example, what are the impact factors influencing the sunroof of the model C 220 in Italy? Therefore, only extrapolative univariate time series approaches are taken into consideration in the following.

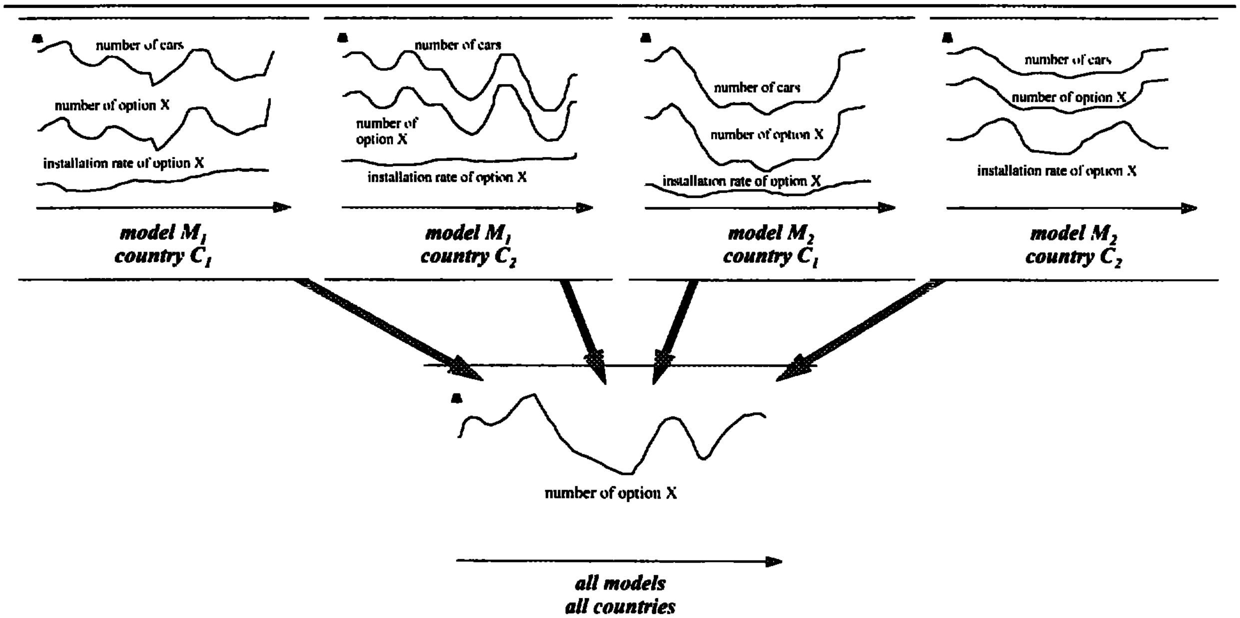

Fig. 3: Installation rates and absolute selling figures of options.

Reviewing extrapolative univariate time series analysis results in quite a lot of different methods. The most important and most popular are listed below (Jarrett, 1991):

- Naive forecast
- Moving averages
- Exponential smoothing
- Time series decomposition
- Adaptive filtering
- The Box-Jenkins methodology (ARIMA modeling)

From the scientific point of view, the most interesting approach of these methods listed above is the ARIMA modeling (Box and Jenkins, 1976). However, there are two very crucial disadvantages: The most important point in ARIMA modeling is model identification. As shown in different forecasting competitions (Makridakis et al., 1984; Makridakis and Hibon, 1979; Schwarze and Weckerle, 1982) even experts differ in choosing the most appropriate ARIMA model. And beyond, there exist forecasting methods producing in average better forecasts than ARIMA modeling does. Besides, as described before, an automatic forecasting system is needed, such that it is not possible to identify manually the appropriate model for each time series. Therefore, ARIMA modeling seems not to be reasonable for the prediction of all time series in our application.

But nevertheless, there exist definitely special types of time series, where ARIMA modeling will be the appropriate forecasting method, i.e. time series with a high autocorrelation. Besides, there exist approaches for identifying automatically the most appropriate model such as Akaike's information criterion (Akaike, 1976; Akaike, 1979). Therefore, the Box-Jenkins methodology will not be the appropriate approach for all time series, but possibly for a selected number of time series.

In the case of adaptive filtering (Kalman, 1960; Kalman and Bucy, 1961), the situation is very similar. It seems not reasonable to use this approach for all time series, but in many cases adaptive filtering will produce good prediction results without any user interaction. Therefore, adaptive filtering will not be used as forecasting technique for all time series, but definitely for the prediction of special types of time series. This result corresponding to the situation of ARIMA modeling is not surprising at all, because the adaptive filtering technique is very similar to ARIMA modeling.

On the one hand, forecasting time series using naive forecasts or moving averages does in a lot of cases not produce good results in comparison to other forecasting approaches. But on the other hand, there exist a lot of time series at Mercedes-Benz, where using naive forecasts or moving averages yields excellent results.

Discussing exponential smoothing (Brown and Meyer, 1961) means to discuss a lot of different forecasting techniques (Gardner, 1985), because there exist constant exponential smoothing approaches without any trend or seasonal parameters as well as there exist different exponential smoothing techniques considering linear and nonlinear trends as well as seasonal effects of the time series. Consequently, the different kinds of exponential smoothing could be depending on the structure of the investigated time series the appropriate approaches for our forecasting system. Generally, the most important and popular types of exponential smoothing approaches are:

- Single exponential smoothing (Brown, 1963)
- Linear exponential smoothing: Brown's one parameter method (Brown, 1963)
- Linear exponential smoothing: Holt's two parameter approach (Holt et al., 1960)
- Winters' three parameter method (Winters, 1960)

Symbolic Machine Learning and Neural Networks. Recently, the attention concerning the task of time series prediction is also focused on the application of neural networks (Weigend and Gershenfeld, 1994; Graf and Nakhaeizadeh, 1993). Although the development of neural networks at early stage was stimulated by modeling of learning process in human brain, the further development of this technology shows a very strong similarity with statistical approaches (Ripley, 1992).

There are some studies which compare neural networks with some statistical procedures like nonlinear regression from a theoretical point of view (Arminger, 1994; Ripley, 1992). However, it should be mentioned that the ability of adaptive learning which characterizes the most of neural networks is not implemented in statistical procedures like regression analysis or ARIMA modeling.

Generally, there are two alternatives to use neural networks for time series prediction:

- On the one hand, it is possible to use neural networks as multivariate forecasting techniques very similar to nonlinear regression. In this case the problem for our application is the same as with multivariate statistical prediction approaches: it is almost impossible to identify the impact factors for the huge number of time series concerning the different options in different countries.
- On the other hand, it is possible to use neural networks as univariate forecasting techniques very similar to statistical extrapolative univariate forecasting approaches. Here the problem is to identify the most appropriate model (Steurer, 1994).

Neural nets based learning algorithms are like the majority of statistical approaches polythetic procedures which consider simultaneously attributes for decision. Because of this reason, they need at least as much training examples as the number of initial attributes. It means, they are not appropriate at all for automatic attribute selection for the cases in which the number of attributes, in comparison to the number of training examples, is too high. This is often the case in dealing with time series data.

The main problem in using neural networks for prediction consists in finding the optimal network architecture. To realize this task, one has to divide the available time series data into training set, test set, and validation set. Regarding the problem of limited number of observations in time series data which is discussed above, dividing the whole series into training set, test set, and validation sets leads to a still smaller training data set, in many circumstances.

Besides neural networks, some of symbolic machine learning algorithms based on ID3-concept (Breiman et al., 1984) can be used to predict the development of time series as well (Merkel and Nakhaeizadeh, 1992).

There are a lot of machine learning algorithms which can handle the classification task. Almost all of these algorithms are, however, not appropriate to handle the prediction task directly. The reason is that in contrast to classification, in prediction the class values are continuous rather than discrete. Exceptions are the ID3-type algorithms CART (Breiman et al., 1984) and Newld (Boswell, 1990) which can handle continuous-valued classes as well. Of course, by discretization of continuous values and building intervals, it is possible to transfer every prediction task to a classification one, but this procedure is connected, normally, with a loss of information.

The algorithms like CART and Newld can handle the continuous valued classes directly, and without discretization. They generate a predictor in the form of a regression tree that can be transformed to production rules. Furthermore, these learning algorithms apply a single attribute at each level of the tree (monothetic) and this is in contrast to the most statistical and neural learning algorithms which consider simultaneously all attributes to make a decision (polythetic).

The main advantage of symbolic machine learning approaches like regression trees is that it is possible very easily to involve other available information in prediction process, for example, by including the background knowledge of experts. Concerning rule generating from the data, it should be mentioned that this property of decision and regression trees is one of the most important advantages of these approaches. Other classification and prediction methods like statistical and neural procedures have not these property that allows considering other sources of information in a very flexible manner. In the case of statistical or neural classification and prediction algorithms, it is very difficult, and in some circumstances impossible at all, to consider such additional information.

However, like other approaches, prediction algorithms based on symbolic machine learning have also some shortcomings. Generally, they can not predict the values beyond the range of training data. Regarding the fact, especially, a lot of time series have an increasing (decreasing) trend component, it can be seen that by using just the raw class values, one can never achieve a predicted value which is outside the range of the class values used for training. But this disadvantage can be avoided by taking differences of the class values as it is the case with the Box-Jenkins approach.

To summarize, neural networks and symbolic machine learning algorithms are from the theoretical point of view very interesting, but they do not seem to be appropriate for the prediction of the huge number of time series in our application. Therefore, the forecast will be done by extrapolative univariate forecasting approaches. The only method of this class not taken into consideration any more is the method of time series decomposition because of the needed user interaction. Therefore, time series decomposition seems definitely not to be an appropriate forecasting method for performing automatic forecasts as it is needed in our application.

The Idea of Combination

After the decision to use different extrapolative univariate time series techniques except for time series decomposition for prediction of options at Mercedes-Benz, the next step will be the detailed selection of an adequate extrapolative univariate prediction model. This selection of the most appropriate forecasting model is a very hard task. In the case of the application presented in this paper, where totally different types of time series have to be forecasted, it is almost impossible to give an answer concerning the question, what generally the most appropriate forecasting model is.

There have been many comparative studies in forecasting. Unfortunately, they do not always agree on which are the best forecasters or on the reasons why one method does well and another does badly. One of the largest studies is that conducted by Makridakis and Hibon (1979) who concluded that simple methods often outperformed sophisticated methods like ARIMA modeling. The largest competition realized so far is that conducted by Makridakis (Makridakis et al., 1984) who concluded with very similar results. While the study of Makridakis and Hibon (1979) concerned 111 time series, the competition of Makridakis (Makridakis et al., 1984) concerned 1001 time series and represents a major research undertaking.

Previous studies by Newbold and Granger (1974) and Reid (1975) concluded that ARIMA modeling will be usually superior if it is appropriate. Generally, ARIMA modeling is regarded as giving extremely accurate forecasts provided the user is expert enough and sufficient computational resources are available, although simple methods are often more than adequate if the stochastic structure of the time series is sufficiently simple (Montgomery and Johnson, 1976; Jarrett, 1991).

In the meantime, many researchers have applied the relatively new techniques of neural networks and machine learning in the context of prediction. There are several recent studies comparing the performance of neural networks, machine learning and classical statistical algorithms using time series data: for example, Rehkugler and Poddig (1990), Schumann and Lohrbach (1992) as well as Graf and Nakhaeizadeh (1993).

All these studies concentrate on finding which method is best, or try to explain why one method is better than another. However, as Bates and Granger (1969) suggest, the point is not to pick the best method, but rather to see if it is possible to combine the available methods so as to get an improved predictor:

> OUR INTEREST is in cases in which two (or more) forecasts have been made of the same event. Typically, the reaction of most statisticians and businessmen when this occurs is to attempt to discover which is the better (or best) forecast; the better forecast is then accepted and used, the other being discarded. Whilst this may have some merit where analysis is the principal objective of the exercise, this is not a wise procedure if the objective is to make as good a forecast as possible, since the discarded forecast nearly always contains some useful independent information.

The general idea of combining different methods is even much older: "In combining the results of these two methods, one can obtain a result whose probability law of error will be more rapidly decreasing" (Laplace, 1818, in Clemen, 1989).

The basic principle of combining forecasts is explained in detail in figure 4. Based on a given database shown in the upper middle of figure 4, forecasting method A discovers the linear trend correctly, but does not detect the seasonality of the given data. In contrast to method A, forecasting procedure B discovers the seasonality but does not detect the linear trend. In conclusion, both methods contain useful information of the given time series, but neither approach A nor method B contains all relevant information. Thus, selecting one single method for prediction and discard the other one means to loose useful and important information. This is avoided by choosing the multistrategy approach of combining both methods to get one common forecast.

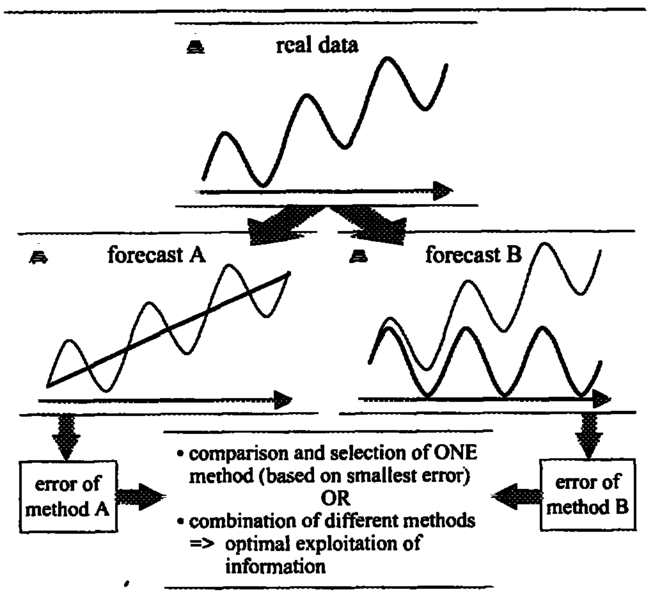

Fig. 4: Combination of forecasts.

This multistrategy approach of combining different forecasting methods is adapted in the considered application at Mercedes-Benz. Instead of selecting one single forecasting procedure, different forecasting techniques will be used in the forecasting system. Each of these different forecasting techniques will produce simultaneously its own forecast for the next 12 (or more) months, that will be combined subsequently in the combination step to one common forecast following Wolpert's general scheme of "Stacked Generalization" (Wolpert, 1992).

Following this approach of combined forecasts, there are three very important questions to answer:

- How many single univariate forecasting methods will be combined ?
- Which single univariate forecasting methods will be combined ?
- What is the appropriate algorithm for the combination ?

The first question to answer is, how many different forecasting approaches should be combined. Theoretically, it can be shown that the larger the number of single forecasting techniques used as input for combination, the better the produced results are (Thiele, 1993). In practice, on the one hand it can be shown, that accuracy will increase if more single forecasts are used as input (Makridakis and Winkler, 1983), but on the other hand, it can be shown as well that the difference of accuracy is not remarkable at all, if there are six, or seven, or eight different single prediction techniques used as input factors for the combination step (Hüttner, 1994). Besides, by using too many single forecasting methods as input for the combination step, there is the danger that the same or almost the same information is taken different times as input for the combination. That means to give a too strong weight to this information and not enough weight to the other independent information. Therefore, it seems reasonable not to use more than six single forecasting procedures as input for the combination step.

Classification of Time Series

The second question is, which forecasting methods should be selected for combination in what situation. As shown in different studies and forecasting competitions (Makridakis et al., 1984; Makridakis and Hibon, 1979; Schwarze and Weckerle, 1982), depending on the characteristics of the time series the different forecasting approaches yield different results. I.e., if there are autoregressive characteristics, ARIMA modeling will perform best, while for other types of time series other approaches will produce the best forecasting results.

Therefore the idea is, to select the six most appropriate single time series approaches depending on the characteristics of the different time series. The goal is to select for every time series the six most appropriate models

time series	attribute							forecasting approach					
no.	stationarity	trend	seasonality	autocorrelation	absolute level of figures	unsystematic fluctuation	...	best	2nd best	3rd best	4th best	5th best	6th best
1	yes	yes	no	no	high	no		E	L	W	A	G	H
2	no	yes	no	no	low	no		R	D	L	G	B	I
3	no	yes	no	no	middle	yes		W	C	L	B	D	G
4	no	no	yes	yes	middle	no		G	W	B	K	M	D
5	yes	no	yes	yes	high	no		F	B	V	E	B	D
6	yes	yes	no	no	low	yes		E	D	F	G	O	K
7	yes	no	yes	yes	low	yes		C	A	U	S	K	L
8	no	no	no	yes	high	no		L	K	S	B	F	J
9	yes	yes	no	no	middle	yes		H	I	D	L	R	C
10	no	no	yes	no	high	yes		A	D	C	G	I	F
...													

Fig. 5: Summary of ex post forecasting results.

and to combine these different forecasts subsequently in the combination step.

For the selection of most appropriate models a randomly selected sample of time series is analyzed and described in detail by using different attributes such as stationarity, trend, seasonality, unsystematic random fluctuation, autocorrelation, absolute level of figures, etc. The values of these attributes are determined using objective tests such as the Dickey-Fuller Test (Dickey and Fuller, 1979) for testing the stationarity and the Durbin-Watson test (Durbin and Watson, 1950, 1951, 1971) for testing autocorrelation.

After this description of all time series in the sample, more than 20 different single univariate prediction procedures as presented before make an ex post forecast for each of those time series. The different forecasting results of these ex post forecasts are compared in detail concerning the prediction quality. This is for analyzing, which prediction procedure performs best, second best, etc. for what type of time series.

As an outcome a table is generated, where the forecasting results are summarized as it is shown in figure 5. In this table the different time series of the selected sample and their characteristic attributes are shown. The single univariate forecasting techniques performing for each time series best, second best, third best etc. are summarized as well.

In best case, it would be sufficient to utilize the table shown in figure 5 to build an automatic decision system, to decide for which type of time series which forecasting techniques performs best, second best, third best, etc. The six best forecasting methods would be selected and used as input for the combination step.

But reality looks differently. For very similar time series with almost the same characteristics the different single prediction techniques perform differently. Therefore, the results of table 5 are not suitable to generate an automatic decision system directly out of it, for what type of time series which forecasting techniques perform best.

Under the assumption that the decision, which forecasting techniques perform best, second best, third best, etc. is a classification task, an appropriate tool to solve such kinds of classification problems is a machine learning algorithm. A classification tool accepts a pattern of data as input, and the output are decision rules or decision trees, which can be used for classifying new and unknown objects (see for more detail Breiman et al., 1984).

Therefore, a symbolic machine learning algorithm is integrated as a generator for an automatic decision system. The ID3-type (Quinlan, 1986, 1987) machine learning algorithm NewId (Boswell, 1990), for example, takes as input a set of examples and generates a decision tree for classification. The product of this machine learning algorithm is a decision tree that can assign an unknown object to one of a specified number of disjoint classes. In our application, such a new unknown object would be a new time series described by its characteristic attributes, and the disjoint classes are the different single forecasting approaches, which are used as input in the combination step.

Using NewId classifications are learned from a set of the attributes, which characterize the different types of time series. For every input of the combination step an extra decision tree is generated. That means that there exist six decision trees: the first decision tree for the first input of the combination step, generated by using all the attributes mentioned in figure 5 and as classification goal the row of best forecasting approaches. The second decision tree is generated by using the same attributes again, but the classification goal is not the row of best but of second best forecasting techniques. The third decision tree is generated by using the row of third best forecasting methods, etc.

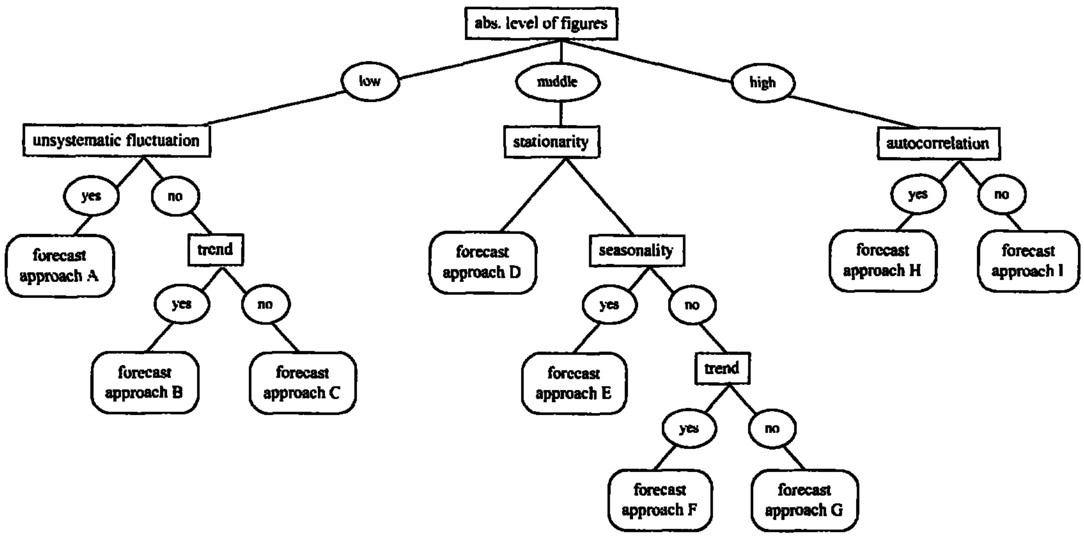

Fig. 6: Decision tree generated by NewId.

Figure 6 shows an example of a decision tree generated by using NewId. Since such a decision tree is generated for each input of the combination step, in sum six decision trees are generated. These six generated decision trees were pruned (Quinlan, 1987) and afterwards used as an automatic decision system, to decide what single forecasting approaches should be used to forecast the different types of time series.

Combining most Promising Approaches

As mentioned above, the third question to answer is, how the combination should be done. Generally, there are two paradigms for combining different forecasting approaches (Henery, 1995):

- On the one hand, it is possible to take the output of a prediction technique and use it as input for another forecast approach. This combination of different models means that only one algorithm is active at the same time (sequential multistrategy forecast).
- On the other hand, the combination of different algorithms at the same time is possible as well. In these approach several algorithms are active at the same time (parallel multistrategy forecast).

In a way, in our application we combine these two alternatives: the first step is kind of parallel multistrategy forecasting, because several forecasting algorithms are active at the same time, when the single prediction methods generate their own forecasts. The second step is kind of sequential multistrategy forecasting, because the output of some prediction methods is used as input for another forecasting procedure.

So far, we consider two different paradigms for the combination algorithm. The first way is to use statistical approaches (Thiele, 1993), the other way is to use neural network approaches (Donaldson and Kamstra, 1996).

From the scientific point of view, the most interesting alternative for the combination would be to use neural networks, i.e. using as learning algorithm the backpropagation algorithm (Werbos, 1974; Rumelhart et al., 1986; Rumelhart and McClelland, 1986). The outputs of the different single forecasting methods selected above are used in this approach as input values for the neural network, where the combination is performed.

But taking into consideration the huge number of time series to forecast, it seems not reasonable to use generally neural networks as combination tool. But nevertheless in some special situations it might be appropriate to use a neural network as forecasting tool in the combination step as it is shown in figure 7.

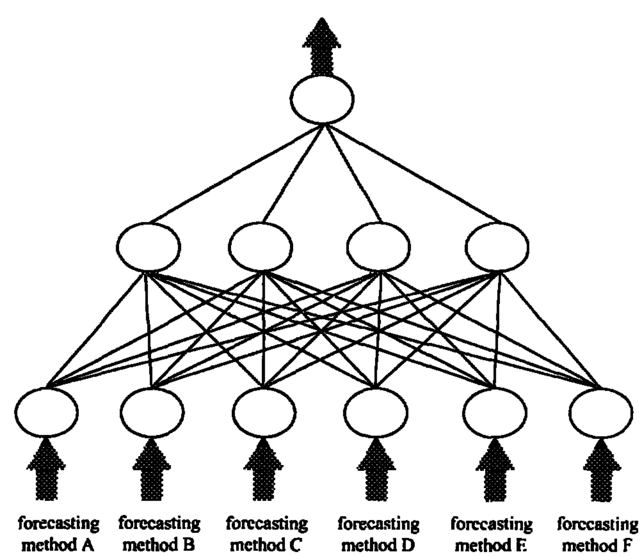

Fig. 7: Neural network as combination tool.

There is one big problem realizing the combination step by using neural networks. As already mentioned above, neural networks normally need a lot of data for training purposes. Because the time series in our application contain in many cases only 30 or 40 values, neural networks in these situations seem not to be the appropriate way for the combination step. But this problem can be solved by using techniques such as the cross-validation method (Lachenbruch and Mickey, 1968; Stone, 1974) and the bootstrap method (Efron, 1983).

The alternative way for running the combination is to use statistical approaches (Thiele, 1993). Generally, there are four classes of statistical combination approaches:

- Unweighted combination: the common forecast is the mean of the different single forecasts.
- Regression methods: the impact of each single forecast for the combination is computed based on a regression analysis (Granger and Ramanathan, 1984).
- Variance-covariance approaches: the impact of each single forecast for the combination is computed based on the error matrix (variance-covariance matrix) of the different methods used in the single prediction step (Bates and Granger, 1969; Clemen, 1989).
- probability-based combination approaches: the impact of each single forecast for the combination is computed based on probabilities, how well every single forecasting method performs (see for more detail Clemen, 1989).

Similar to the single forecasting approaches the best alternative of doing the combination depends on the characteristics of the time series. Therefore the same way is chosen as it is done for the selection of the best single forecasting methods. Again the machine learning algorithm Newld is used for generating a decision tree. This decision tree is used for an automatic selection, in which situation which way for combination seems most appropriate.

In many cases, already the simple mean of the different single forecasting procedures yields much better results than the best single forecast technique does. This confirm Kang (1986) ("A simple average is shown to be the best technique to use in practice.") and Clemen (1989) ("In many studies, the simple average of the individual forecasts has performed best or almost best".). But there are other situations as well, where it is appropriate to combine the forecasts of the single forecasting approaches by using more sophisticated techniques.

Theoretically, the regression and the variance-covariance approach yield identical results (Thiele, 1993), but caused by estimation errors in estimating the different parameters, in practice the regression method and the variance-covariance method produce different forecasts.

Sometimes, best forecasting results are produced by using probability-based approaches for running the combination step. The most interesting task using this alternative is the estimation of a priory probabilities, which are needed for running the combination of the different single forecasts (Clemen, 1989).

Evaluation

For evaluation purposes, there are two different parts of evaluation: The first part of evaluation is done by taking the DGOR (Deutsche Gesellschaft für Operations Research = German Society of Operations Research) forecasting competition of 1982 (Schwarze and Weckerle, 1982) as a general benchmark. The second and concerning the real application much more interesting part is to test the new multistrategy approach in comparison to the actual forecasting procedure, which is basically a weighted moving average method.

DGOR Forecasting Competition

The first part of evaluation is done by taking the DGOR forecasting competition of 1982 (Schwarze and Weckerle, 1982) as a general benchmark.

This competition was organized in 1982 by the forecasting group of the German Society of Operations Research and the task is to forecast 15 different time series of monthly data. All time series are real data concerning selling figures, turnover figures, etc. The longest time series contains 204 values, the shortest one contains 48 values.

Generally, there are two parts of the competition, where each time the task is to forecast the above mentioned 15 time series:

- In the first part of the competition the task is to forecast once 12 periods forecasting horizon.
- In the second part of the competition the task is to forecast 12 times 1 period forecasting horizon.

As an evaluation criteria for performance of the different forecasts Theil's U (Theil, 1971) is used.

For forecasting once 12 periods forecasting horizon, Theil's U is calculated as shown in (1) and for forecasting 12 times 1 period forecasting horizon, U is calculated as shown in (2):

$$U_1 = \sqrt{\frac{\sum_{t=t_0+1}^{t_0+T}(x_t - \tilde{x}_t)^2}{\sum_{t=t_0+1}^{t_0+T}(x_t - x_{t_0})^2}} \quad (1) \qquad U_2 = \sqrt{\frac{\sum_{t=t_0+1}^{t_0+T}(x_t - \tilde{x}_t)^2}{\sum_{t=t_0+1}^{t_0+T}(x_t - x_{t-1})^2}} \quad (2)$$

- x_t : real value of month t.
- $\tilde{x}_t$: forecasted value for month t.
- t_0: actual month, where the forecast is done.
- T: total forecasting horizon (in our application: 12 months).

As shown in (1) and (2) U is calculated as square root of the quotient of the squared forecasting error divided by the squared forecasting error of the naive forecast.

Therefore, the most important advantage of using Theil's U is the fact that U automatically gives an answer concerning the question, whether it is worth at all using a more or less complex forecasting approach or whether it is sufficient to perform a prediction using the naive forecast:

- $U = 0$: perfect forecast, it is worthwhile to forecast.
- $0 < U < 1$: forecast better than naive forecast, it is worthwhile to do the forecast.
- $U = 1$: forecasting error equal to naive forecast, it is not worthwhile to do the forecast.
- $U > 1$: forecast worse than naive forecast, it is not worthwhile to do the forecast.

Using Theil's U as an evaluation criteria for performance of our forecasting system and taking the mentioned DGOR forecasting competition as a general benchmark, the approach presented yields very good results:

- In the first part of the competition (1 time 12 periods forecasting horizon) our method performs U_1=0.69, which is the second best result (best approach: U_1^{best}=0.67).
- In the second part of the competition (12 times 1 period forecasting horizon) our method performs U_2=0.58, which is also the second best result (best approach: U_2^{best}=0.55).

Note that approaches reaching better prediction results (i.e. lower U value) need manual fine tuning and produced better results after being tuned based on each single time series, whereas our combined approach presented in this paper works without any manual interaction and standard parameter values, but completely automatically.

Mercedes-Benz Data

The second and concerning the real application much more interesting part of evaluation is to test the new multistrategy approach in comparison to the actual forecasting procedure, which is basically a weighted moving average method.

Based on 200 randomly selected time series that are different to those time series used for performing the classification task in the previous section the new approach is evaluated by making ex post forecasts of the last 12 months. For this purpose similar to the evaluation using DGOR data, there are again two steps of evaluation, where each time the task is to forecast the above mentioned 200 time series of Mercedes-Benz data.

- In the first step the task is to forecast once 12 periods forecasting horizon.
- In the second step the task is to forecast 12 times 1 period forecasting horizon.

In earlier tests (Friedrich et al., 1995) it was already shown, that the actual weighted moving average forecasting method produces better results than the naive forecast does ($U < 1$). Therefore, it seems not reasonable at all to compare the new combined forecasting approach with the naive forecast as it is done by using Theil's U such as in the previous subsection.

Instead of using Theil's genuine U as an evaluation criteria a modified version of Theil's U as shown in (3) and (4) is used as evaluation criteria. In comparison to Theil's genuine U the modified U^* is calculated as square root of the quotient of the squared forecasting error not divided by the squared forecasting error of the naive forecast but divided by the squared forecasting error of the actual weighted moving average forecasting method. For forecasting once 12 periods forecasting horizon, the modified U^* is calculated as shown in (3) and for forecasting 12 times 1 period forecasting horizon, the modified U^* is calculated as shown in (4):

$$U_1^* = \sqrt{\frac{\sum_{t=t_0-1}^{t_0+T}(x_t-\tilde{x}_t)^2}{\sum_{t=t_0-1}^{t_0+T}(x_t-\hat{x}_t)^2}} \quad (3) \qquad U_2^* = \sqrt{\frac{\sum_{t=t_0+1}^{t_0-T}(x_t-\tilde{x}_t)^2}{\sum_{t=t_0+1}^{t_0+T}(x_t-\hat{x}_t)^2}} \quad (4)$$

- x_t : real value of month t.
- $\tilde{x}_t$: forecasted value for month t using the new combined forecasting approach.
- $\hat{x}_t$: forecasted value for month t using the actual weighted moving average forecasting approach.
- t_0: actual month, where the forecast is done.
- T: total forecasting horizon (in our application: 12 months).

Similar to Theil's genuine U using Theil's modified U^* automatically gives an answer concerning the question, whether it is worth at all using the new approach or whether it is sufficient to perform a prediction using the actual weighted moving average forecasting approach:

- $U^* = 0$: perfect forecast, it is worthwhile to use the new approach.
- $0 < U^* < 1$: combined forecast better than actual forecast, it is worthwhile to use the new method.
- $U^* = 1$: forecasting error of combined forecast equal to that of actual method, it is not worthwhile to use the new approach.
- $U^* > 1$: combined forecast worse than actual method, it is not worthwhile to use the new approach.

Using the modified U^* as an evaluation criteria for performance of our forecasting system and taking the actual forecasting method as a benchmark, the combined approach presented yields very good results:

- In the first step (1 time 12 periods forecasting horizon) our new approach performs U^*_1=0.67.
- In the second step (12 times 1 period forecasting horizon) our new method performs U^*_2=0.68.

Obviously, the combined forecasting approach yields much better results than the actual method does. But due to the fact that neither Theil's genuine U nor the modified U^* are linear measures, it is very difficult to say exactly (i.e. in percent) how much better the new combined forecasting approach is in comparison to the actual method.

In summary, in comparison to the actual forecasting technique, our new combined approach yields much better results. Due to the fact that the results are confidential, they cannot be reported in more detail.

References

Akaike, H. 1976. Canonical correlation analysis of time series and the use of an information criterion. In: R.K. Mehra, and D.G. Lainiotis (eds.): System Identification, Academic Press, New York.

Akaike, H. 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting: *Biometrika 66: 237-242.*

Arminger, G. 1994. Ökonometrische Schätzmethoden für neuronale Netze. University of Wuppertal.

Bates, J.M., and Granger, C.W.J. 1969. The combination of forecasts: *Operations Research Quarterly 20: 451-468.*

Box, G.E.P., and Jenkins, G.M. 1976. Time series analysis, forecasting and control. Holden-Day, San Francisco.

Boswell, R. 1990. Manual for NewId. The Turing Institute.

Brändli, N., and Dumschat, R. 1995. Regelbasierte Stücklisten im Anwendungsdatenmodell ISO 10303-214: *Produktdaten Journal 1: 10-12.*

Brändli, N., Käfer, W., and Malle, B. 1994. Enhancing Product Documentation with new Information Technology Systems Based On Step: Proceedings of the 27th ISATA: Mechatronics and Supercomputing Applications in the Transportation Industries, 429-436.

Breiman, L., FRiedmann, J.H., Olshen, R.A., and Stone, C.J. 1984. Classification and regression trees. Wadsworth and Brooks, Monterey.

Brown, R.G. 1963. Smoothing, forecasting and prediction of discrete time series. Prentice Hall, Englewood Cliffs.

Brown, R.G., and Meyer, R.F. 1961. The fundamental theorem of exponential smoothing: *Operations Research 9: 673-685.*

Clemen, R.T. 1989. Combining forecasts: a review and annoted bibliography: *International Journal of Forecasting 5: 559-583.*

Dichtl, E., Raffée, H., Beeskow, W., and Kögelmayr, H.G. 1983. Faktisches Bestellverhalten als Grundlage einer optimalen Ausstattungspolitik bei Pkw-Modellen: *ZFBF 35 (2): 173-196.*

Dickey, D.A., and Fuller, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root: *Journal of the American Statistical Association 74: 427-431.*

Donaldson, R.G., and Kamstra, M. 1996. Forecast combining with neural networks: *Journal of Forecasting 15: 49-61.*

Durbin, J., and Watson, C.S. 1950. Testing for the serial correlation in least squares regression (I): *Biometrika 37: 409-428.*

Durbin, J., and Watson, C.S. 1951. Testing for the serial correlation in least squares regression (II): *Biometrika 38: 159-178.*

Durbin, J., and Watson, C.S. 1971. Testing for the serial correlation in least squares regression (III): *Biometrika 58: 1-19.*

Efron, B. 1983. Estimating the error rate of a prediction rule: improvements on cross-validation: *Journal of the American Statistical Association 78: 316-331.*

Friedrich, L., Hirzel, J., Kirchner, S., and Ohl, S. 1996. Ein neues Prognoseverfahren in der Automobilindustrie. Forthcoming.

Friedrich, L., and Kirchner, S. 1995. Das Prognose-Modell SAIPRO, Forschungsberichte 1-5: Technical Reports, Daimler-Benz AG, Berlin, Ulm.

Friedrich, L., Fooken, J., Hoffmann, S., Ludwig, R., Müller, E., Nehls, S., and Schütz, F. 1995. Forschungsbericht: Analyse und Prognose von Kundenauftragsdaten, Dept. of Mathematics and Physics, Technische Fachhochschule Berlin.

Gardner, E.S. Jr. 1985. Exponential smoothing: state of the art: *Journal of Forecasting 4: 1-28.*

Graf, J., and Nakhaeizadeh, G 1993. Application of neural networks and symbolic machine learning algorithms to predicting stock prices. In: V.L. Plantamura, B. Soucek, and G. Vissagio (eds.): Logistic and learning for quality software management and manufacturing. John Wiley & Sons, New York.

Granger, C.W.J., and Ramanathan, R. 1984. Improved methods of combining forecasts: *Journal of Forecasting 3: 197-204.*

Henery, R.J. 1995. combining forecasting procedures. In: Y. Kodratoff, G. Nakhaeizadeh, and C. Taylor (eds.): Workshop Notes of the Mlnet Familiarization Workshop *Statistics, Machine Learning and Knowledge Discovery in Databases.*

Hirzel, J. 1995. Programmplanung für variantenreiche Automobile im Wandel vom Verkäufer- zum Käufermarkt. In: G. Bol, T. Christ, and B. Suchanek (eds.): Das Potential der Lean-Technik in der Kraftfahrzeugwirtschaft: Aktuelle Entwicklungen der Material-Logistik in Theorie und Praxis.

Hirzel, J., and Ohl, S. 1995. Forecasting of Extras and Options in the Car Industry. In: P. Kleinschmidt (ed.): Operations Research Proceedings 1995, 409-410. Springer, Heidelberg.

Holt, C., Modigliani, F., Muth, J.F., and Simon, H.A. 1960. Planning production, inventories and work force. Prentice Hall, Englewood Cliffs.

Hüttner, M. 1994. Vergleich und Auswahl von Prognoseverfahren für betriebswirtschaftliche Zwecke. In P. Mertens (ed.): Prognoserechnung. 5., neu bearbeitete und erweiterte Auflage, Physika, Heidelberg.

Jarrett, J. 1991. Business forecasting methods. Second Edition, Basil Blackswell, Oxford.

Conclusion

Summary

In conclusion, the task was to forecast cumulative time series of single options in the car industry, that do - at first glance - not contain any systematic pattern like trend or seasonality at all. By utilizing descriptive statistics, we discovered that the installation rate of single options for different models in different countries is more or less systematic and structured. Therefore, the forecast is done on installation rates, that are multiplied by the planned number of cars produced per month of each model for each country. Due to the huge number of time series to predict, automatic forecasting approaches without any manual user interaction are needed.

Because it is very difficult to identify external impact factors, no causal approaches like regression analysis were taken into consideration. Machine learning approaches and neural nets for running the prediction seem not to be appropriate as well. Therefore, univariate extrapolative prediction methods were chosen for running the forecast. But instead of selecting one single forecasting technique such as ARIMA modeling more than 20 different univariate forecasting approaches are tested for a randomly selected sample of time series. Depending on the characteristics of the time series, an automatic decision model was generated by using the machine learning algorithm NewId.

Using this decision model, for every time series the six most appropriate forecasting approaches are selected, and subsequently these six single forecasts are combined to get one common forecast. Generally, for performing this combination neural networks seem reasonable as well as statistical approaches. Which alternative is used for the combination step is again decided by a decision tree, that was generated again by using the NewId algorithm.

Selecting the most appropriate forecasting approaches depending on the characteristics of the time series and selecting again the most appropriate combination method for the combination seems to guarantee an automatic adaptation of the forecasting system to the time series.

In summary, in this paper a two step multistrategy forecasting approach is presented, that integrates machine learning techniques like decision trees for generating the automatic decision systems as well as several well known statistical univariate time series approaches such as ARIMA modeling, adaptive filtering, and exponential smoothing. Besides, multivariate statistical approaches such as linear regression are used in this approach as well as neural networks for performing the combination step. An overview of the total system is given in figure 8.

Future Work

Currently, improvement of single forecasting methods by automatic adaptation of parameters and improvement of the decision trees by increasing the sample of time series is under consideration. Besides, testing alternative combination approaches such as using other types of neural networks and other types of machine learning algorithms like Quinlan's M5 (Quinlan, 1992) for running the combination step are under consideration as well.

In addition, recently a completely new single forecasting approach called SAIPRO was developed (Friedrich and Kirchner, 1995; Friedrich et al., 1996), that will be integrated in the combined forecasting system next.

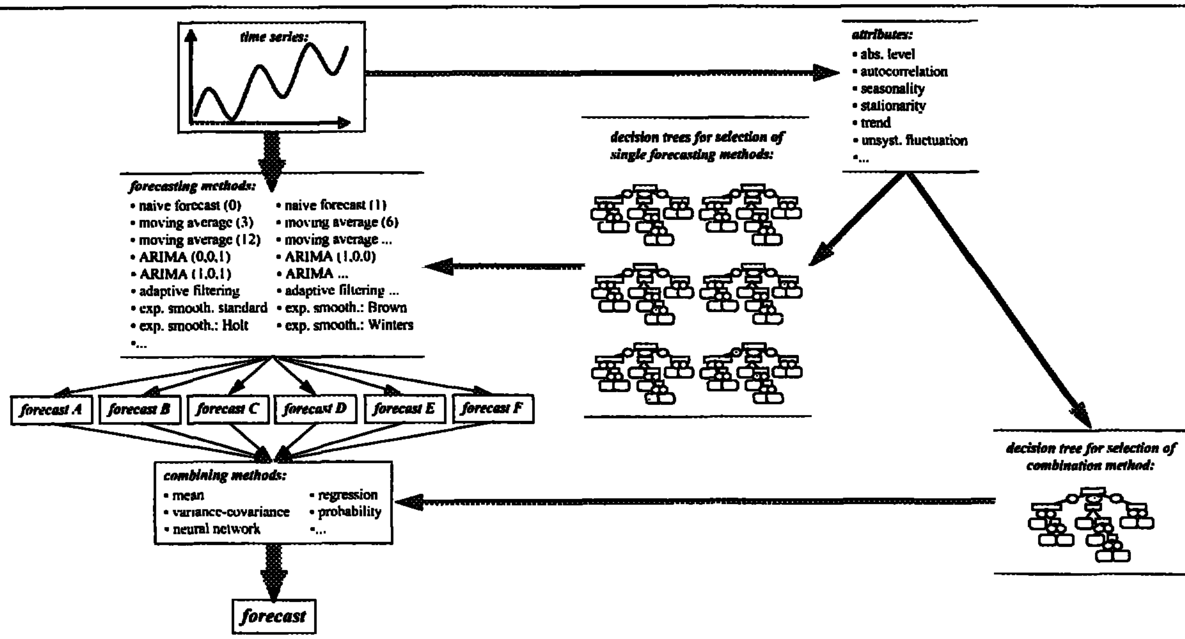

Fig. 8: Overview of the multistrategy forecasting system.

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems: *Journal of Basic Engineering 82: 35-44.*

Kalman, R.E., and Bucy, R.S. 1961. New results in linear filtering and prediction theory: *Journal of Basic Engineering 83: 95-107.*

Kang, H. 1986. Unstable weights in the combination of forecasts: *Management Science 32: 683-695.*

Lachenbruch, P., and Mickey, R. 1968. Estimation of error rates in discriminant analysis: *Technometrics 10: 1-11.*

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen E., and Winkler, R. 1984. The forecasting accuracy of major time series. John Wiley & Sons, New York.

Makridakis, S. and Hibon, M. 1979. Accuracy of forecasting: an empirical investigation (with discussion): *Journal of the Royal Statistical Society A 142: 97-145.*

Makridakis, S. and Winkler, R.L. 1983. Averages of forecasts: some empirical results: *Management Science 29: 987-995.*

Merkel, A., and Nakhaeizadeh, G 1992. Application of artificial intelligence methods to prediction of financial time series: *Operations Research 91: 557-559.*

Montgomery, D.C. and Johnson, L.A.. 1976. Forecasting and time series analysis. McGraw-Hill, New York.

Newbold, P., and Granger, C.W.J. 1974. Experience with forecasting univariate time series and the combination of forecasts: *Journal of the Royal Statistical Society A 137: 131-165.*

Ohl, S. 1995. Analysis of the behavior of new car buyers. In: H.H. Bock and W. Polasek (eds.): Data analysis and information systems, statistical and conceptual approaches, studies in classification, data analysis, and knowledge organization, 197-207. Springer, Heidelberg.

Ohl, S. 1996. Forecasting virtual customer orders vs. forecasting combinations of options. Forthcoming.

Quinlan, R. 1986. Induction of decision trees: *Machine Learning 1: 81-106.*

Quinlan, R. 1987. Simplifying decision trees: *International Journal of Man-machine Studies 27: 221-234.*

Quinlan, R. 1992. Learning with continuous classes. In: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence. World Scientific, Singapore.

Rehkugler, H., and Poddig, T 1990. Statistische Methoden vs. künstliche Neuronale Netzwerke zur Aktienkurs-prognose - Eine vergleichende Studie. Discussion paper, University of Bamberg.

Reid, D.J. 1975. A review of short-term projection techniques. In: Gordon, H. (ed.): Practical aspects of forecasting, Operational Research Society, London.

Ripley, B.D. 1992. Statistical aspects of neural networks. University of Oxford.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. 1986. Learning representations by back-propagating errors: *Nature 323: 533-536.*

Rumelhart, D.E., and McClelland, J.L. 1986. Parallel distributed processing. Vol. I & II. MIT Press, Cambridge.

Schwarze, J., and Weckerle, J. 1982. Prognose-verfahren im Vergleich. Technical report. University of Braunschweig.

Schumann, M., and Lohrbach, T 1992. Artificial neural networks and ARIMA models within the field of stock market prediction - a comparison. In: Preliminary Proceedings, BANKAI Workshop on Adaptive Intelligent Systems, Belgium, S.W.I.F.T.

Stegemann, B. 1995. Dampf in allen Klassen: *Auto Motor Sport 3: 8-13.*

Steurer, E. 1994. Prognose von 15 Zeitreihen der DGOR mit Neuronalen Netzen und Vergleich mit den Methoden des Verfahrensvergleichs der DGOR aus dem Jahre 1982: Technical Report, Daimler-Benz AG, Ulm.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions: *Journal of the Royal Statistical Society 36: 111-133.*

Theil, H. 1971. Applied Economic Forecasting. North-Holland, Amsterdam.

Thiele, J. 1993. Kombination von Prognosen. Physika, Heidelberg.

Weigend, A.S., and Gershenfeld, N.A. 1994. Time series prediction: Forecasting the future and understanding the past. Addison-Wesley, Reading.

Werbos, P. 1974. Beyond regression. Ph.D. Thesis, Harvard University.

Winters, P.R. 1960. Forecasting sales by exponentially weighted moving averages: *Management Science 6: 324-342.*

Wolpert, D.H. 1992. Stacked generalization. *Neural Networks, 5: 241-259.*