Evolving Multilevel Forecast Combination Models - An Experimental Study

Silvia Riedel*,** and Bogdan Gabrys**
* Lufthansa Systems Berlin GmbH, Fritschstrasse 27-28, 10585 Berlin Germany
Phone: +49-30-34007141, Fax: +49-30-34007100
email: silvia.riedel@lhsystems.com
** Computational Intelligence Research Group, School of Design, Engineering & Computing
Bournemouth University, Poole House, Talbot Campus
Poole, BH12 5BB United Kingdom
Phone: +44-1202-595298, Fax: +44-1202-595314
email: bgabrys@bournemouth.ac.uk

ABSTRACT: This paper provides a description and experimental comparison of different forecast combination techniques for the application of Revenue Management forecasting for Airlines. In order to benefit from the advantages of forecasts predicting seasonal demand using different forecast models on different aggregation levels and to reduce the risks of high noise terms on low level predictions and overgeneralization on higher levels, various approaches based on combination of many predictions are presented and experimentally compared. We propose to evolve combination structures dynamically using Evolutionary Computing approaches. The evolved structures are not only able to generate predictions representing well balanced and stable fusions of methods and levels, they are also characterised by high adaptive capabilities. The focus on different levels or methods of forecasting may change as well as the complexity of the combination structure depending on changes in parts of the input data space in different data aggregation levels. Significant forecast improvements have been obtained when using the proposed dynamic multilevel structures.

KEYWORDS: Forecast Combination, Adaptive Forecasting, Genetic Programming Airline, Revenue Management

1 INTRODUCTION

A typical approach to building a forecast model consists of a phase of data analysis, determination of a level of forecasting and history building, model creation, determination of appropriate preprocessing, parameter calibration and validation. The model, once built, is updated on a regular basis by model rebuilding or updating the history based on the most recent data. But even if the calibration has been done well, it is likely that the real relationship between given inputs and the values to predict is so complex that it can not be modelled perfectly. A common problem is also the detection of appropriate input data aggregation levels on the basis of which forecasts are generated. If the aggregation level is too low, there is a risk of being able to observe only a small number of historical data containing very high levels of random noise and not taking into account important information. If, on the other hand, the data aggregation level is too high, there is a risk that important data characteristics are not represented appropriately. This problem is becoming even more relevant if the underlying processes and data change over time and the chosen settings are not optimal any more.

Considerable literature has accumulated regarding the combination of forecasts (for overviews see e.g. [1] or [2]) concluding that the forecast accuracy can be substantially improved through the combination of multiple individual forecasts. In the presented study different approaches for combining predictions are used in order to benefit from the advantages of forecasts predicting on different levels of the input data space and at the same time to reduce the risks highlighted earlier. Combinations using all forecasts in one combining process are compared to different multistep approaches.

We propose to evolve combination structures dynamically using Evolutionary Computing approaches. The evolved structures are not only able to generate predictions representing well balanced and stable fusions of methods and levels, they are also characterised by high adaptive capabilities. The focus on different levels or methods of forecasting may change as well as the complexity of the combination structure depending on changes in parts of the input data space in the different data aggregation levels.

We first discuss the problem in giving an introduction to combining approaches and a motivation of why we decided to use multilevel forecasts on decomposed data. Then the use of static multilevel combination structures is shortly described. The focus in this study is put on an investigation of how genetic programming approaches can be used in order to evolve dynamic multilevel combination structures. We finally discuss experimental results which show a clear forecast improvement while using dynamic multilevel combination strategies for the Revenue Management application of demand forecasting for airlines with special focus on the component representing seasonal behavior.

1.1 COMBINATION APPROACHES

Forecast combination approaches are today a scientifically acknowledged procedure to model complex functional relationships in producing not one optimal forecast, but a number of forecasts $\widehat{x}^m \in \mathcal{R}$ and combining them for the final prediction $\widehat{x} \in \mathcal{R}$. The existing combination approaches differ in the description of the functional relationship $f : \mathcal{R}^m \rightarrow \mathcal{R}$ which represents the fusion process. In linear combination models the relationship is defined as a simple weighted linear sum of the individual forecasts:

$$\widehat{x} = \sum_m w_m \widehat{x}^m, \; w_m \in \mathcal{R} \; \forall m. \tag{1}$$

Individual forecast performance is taken into account to calculate the weights (for details concerning models see e.g.[1] and [2]). A more complex and flexible group of combination models are nonlinear combination models. In this group, mostly application specific approaches differ in the selection of external input information as well as in the general approach. Typical nonlinear approaches include neural networks [3], (fuzzy) expert systems [4] and functional approaches.

It has been shown theoretically and experimentally (see e.g. [5]) that the best results can be achieved if different individual forecasts are divers in the sense that they are able to provide diverse knowledge based on different input information in terms of different available sources of information, different preprocessing or history pools, different functional or stochastic approaches or different parametrisation of the models. So combination can also be used in order to profit from different parameter settings or different views of historical data. It is even possible to use it to calibrate preprocessing like signal stabilisation and filtering procedures. Note that the advantage here is that the combined forecast is profiting not only from one often very good solution like one optimised parameter set, but from a whole set and additionally it is an adaptive procedure meaning that, e.g., a parameter set would automatically adapt to new situations.

For our application of demand forecasting for airlines it was not possible to significantly outperform the individual forecasts using the common combination models directly on different predictions made for the future demand. A reason for the only small improvements can be found in the high covariance values between the individual forecasts. These result from high noise terms in the data because of an extremely fine forecast level and forecast models which take into account the same demand influencing factors and therefore produce results which are not very diverse. The process of learning the combining weights is over-interpreting small changes in the forecasts, which leads to instabilities and data overfitting. The observed behaviour is a common problem occurring with the use of combination models. For similar experiences described in the literature see e.g. [5].

In order to improve the situation we followed the strategy of decomposition and multilevel forecasting. A motivation for this decision is given in the next subsections.

1.2 DECOMPOSITION

Data is often influenced by a whole set of impacts, which are (hopefully) rather independent of each other. Generally not all of these impacts are known or modelled. That is why predictions, even if generated using different models, often produce errors which are correlated to the not modelled impacts and which are therefore also correlated among each other. This leads to the problems for the combination process mentioned above.

The approach of decomposition offers not only a solution in order to avoid the combination of highly correlated forecasts (see e.g. [6]), it also brings a lot of other advantages like the use of models and parameter sets which are specialised to the characteristics of a specific component concerning e.g. its structure, dependencies on input information, stability and noise level.

1.3 DETERMINATION OF APPROPRIATE FORECAST LEVELS

Real world forecasting problems are often not related to an isolated prediction, but to the task of predicting the future situation in an application defined space. For the application of revenue management forecasts this means a whole network of flights, routings or itineraries which again are containing different fareclasses, point of sales and other variables. A crucial problem is to determine appropriate levels on which forecasts should be produced, on which the models are calibrated or structural characteristics are determined. If a level is chosen too fine, there is a high risk of undesirable small numbers predictions in connection with large noise terms in the data as observed in the experiments described above. Important characteristics or relations may not or only badly be detected, which leads to unstable or structurally poor forecasts. If on the other hand the chosen level is too general, important characteristics related to special parts in the input space may be ignored.

The task is getting even more difficult if the data is composed of different independent components. So it is likely that seasonal behavior should be modelled on another level than a schedule change. In practice the level of calculation is often determined based on static test data. This may become a problem when the behavior changes over time. So it can be for instance good to learn seasonal behavior for a certain booking class on the class level, but if suddenly the amount of bookings in this class is significantly reduced because a new competitor is on the market, the seasonal behavior should better be learned on a higher level (e.g. per compartment) in the future.

Automatic approaches to solve the problem should ideally have the following characteristics:

- The forecasts should be able to profit from specialists knowledge (fine level forecasts).
- The forecasts should be able to profit from general structural knowledge (higher levels).
- Fine level noise influence should be reduced.
- The selection of the levels should be adapted automatically to changes in the data.

In [7] we have proposed to extend the use of combining models to generate forecasts on decomposed data, which use all available information, meaning also different input data representations $\hat{x}_l^m \in \mathcal{R}$ from different levels l, and which are therefore also able to adapt automatically to new situations present at different levels.

2 STATIC STRUCTURES OF MULTILEVEL COMBINATIONS

2.1 Flat Structures of Multilevel Combinations

When applying just one combination procedure on multilevel forecasts (see figure 1) a crucial problem is that the potential number of forecasts to combine increases significantly. The experience described in the literature showed as well as our own experiments that the use of a big number of forecasts can lead to solutions which are not able any more to identify relevant information. There also exists a much higher risk of instabilities based on numerical problems because of noisy data or pour generalisation ability due to data overfitting problems (see e.g. [8] and [2]).

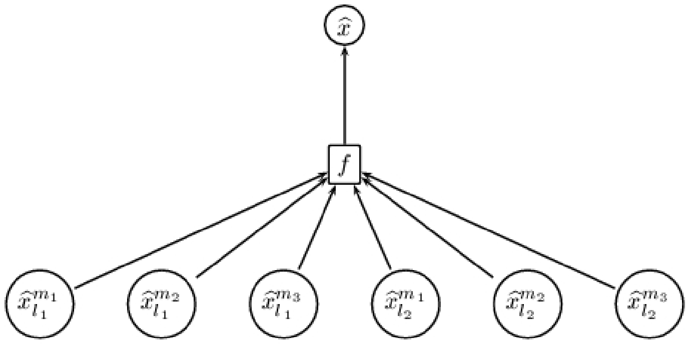

Figure 1: Flat combination of 6 forecasts produced using three functional approaches m_1 to m_3 at two levels l_1 and l_2.

We could achieve forecast improvements if we did the ”correct selection” of not more than five multilevel input forecasts for the combination, especially if these were connected with quite specialised nonlinear combination models. The inconvenience here is that the choice of the ideal set of input forecasts demands a high level of special knowledge about the application and that the approach is restricted in the sense that the choice is not adaptive to changes occurring in the future.

The observed effects correspond to what can be observed in the process of decision finding in the society. If to many people are included into the process of discussion, the resulting decisions are often not optimal, but compromises tending to satisfy everyone. Two kinds of alternatives have evolved in society: to reduce the number of people to a set of specialists or to build up hierarchies and to generate the decisions in multistep procedures.

2.2 HIERARCHICAL STRUCTURES OF MULTILEVEL COMBINATIONS

As just mentioned, a possibility to handle the problem of the big set of potential input forecasts is not to combine all forecasts but to use multistep approaches combining a number of forecasts and giving the result to a next combination step. We have generated different intuitive hierarchical structures based on diversity aspects of potential input forecasts. The structures have been generated in a manner that one step is generally combining predictions which are diverse concerning one diversity criteria (e.g. the use of a functional approach to generate the prediction), but comparable concerning the others (e.g. concerning the level of history building and parametrisation). Figure 2 illustrates one of the most successful structures. It is a hierarchical top down approach meaning that first the forecasts of higher levels are combined and then the combination of the forecasts of lower levels, but using also the combined forecast from the higher level, is performed.

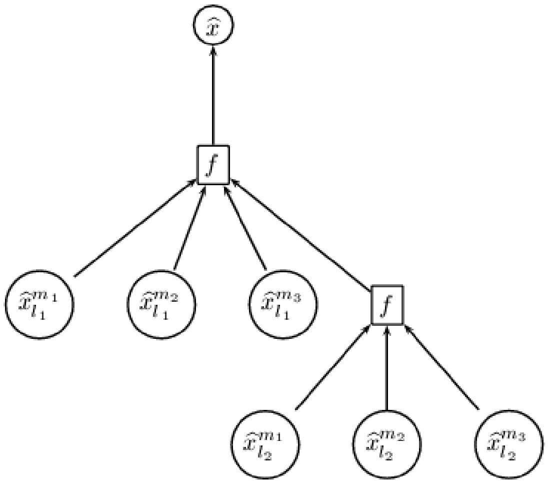

Figure 2: Hierarchical multilevel forecast combination using three functional approaches m_1 to m_3 at two levels l_1 and l_2. First the higher level predictions $\widehat{x}_{l_2}^{m_1}$ to $\widehat{x}_{l_2}^{m_3}$ are combined, then the result is passed to the low level combination.

3 DYNAMIC STRUCTURES OF MULTILEVEL COMBINATIONS

The result obtained using the intuitive structures have been quite diverse. Some of the structures produced very encouraging results, others did not outperform the flat structures or did even worse. But even if the good results show that the use of multistep structures may be a way to overcome the problems described in the previous section, the approach of using predefined intuitive structures is limited in the sense that it needs a lot of expert knowledge in order to identify promising ones. This task is getting even harder by the fact that a lot of decisions have to be made in advance like the choice of the number of steps and the input forecasts to include. Potential structures, once identified, then have to be verified by experiments using try and error principles. And as the fixed structures contain only limited adaptive capabilities, they would have to be rebuilt on a regular basis.

The best structures do not necessarily need to be the intuitive ones. Often solutions found in nature do not follow our intuition. All these reasons motivate to search for dynamic approaches generating and adapting structures automatically. The generated optimal structures need to be able to work well in a changing environment. Evolutionary computation offers common algorithms to solve such kind of problems. It simulates evolution in applying optimisation algorithms which iteratively improve the quality of solutions until an optimal, or at least high quality solution is found. As evolution continues over time the iterative process generates solutions which have proven to be flexible in a changing environment by having survived different generations.

3.1 SPECIFICATION OF THE OPTIMISATION PROBLEM

In order to be able to use evolutionary computation to generate combination structures dynamically, we first have to describe our problem as an optimisation problem including a description of general conditions of the problem together with restrictions and the criterium to be optimized. This will be described in this subsection. We start with a discussion of how to come to a restricted set of input forecasts, continue with a description of our search space as a set of valid combination structures and finally provide a criterium to be optimized which is based on forecast quality of the resulting combined predictions.

3.1.1 *Input forecast selection*

As individual forecasts can differ with regard to the use of

- different available sources of information,
- different functional or stochastic approaches,
- different parametrization or the
- different levels of decomposition,

the space of potential input forecasts to be combined can get very large. That is why it is generally not possible to include all of them into a combination process. There are different options of how to handle this problem.

The first option is to chose a representative set randomly or by expert selection. The set should by chosen small in a manner that it covers well the complete space of potential input forecasts. It can be expected that one does not loose to much relevant information in including only those forecasts into the evolution, because forecasts differing in only one dimension, e.g. only by small parameter changes, are often highly correlated so that the information loss is not critical.

The second option is also to start with a subset of input forecasts, but to extend or change this set during the evolution process. If it is learned that certain forecasts are especially relevant, it may be useful to include other forecasts which have similar characteristics.

3.1.2 *Description of the search space*

The search space is the space of valid combination structures which will be now described. A valid multistep combination structure contains one or more combination procedures. Each combination procedure is related to a combination step i, different steps are connected by the fact that the input forecasts of a combination procedure of step $i > 1$ correspond to result forecasts of combination procedures of step $i - 1$. The set of input forecasts of a combination procedure at step 1 is a subset of the set of input forecasts.

Figure 3.1.2 shows a combination process as a tree. The calculation of the combination represented by the tree is carried out from the leaves to the root. The leafs represent the input predictions used in the first combination procedures of step 1. All nodes represent combination procedures using one fixed or different combination methods and combining the predictions passed to them from the lower subtree. The root is the final combination generating the combined forecast.

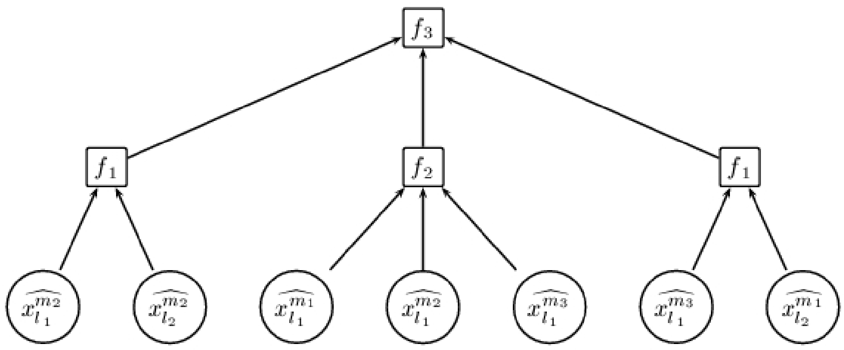

Figure 3: An example for a combination structure as a tree combining multilevel forecasts generated using using three functional approaches m_1 to m_3 at two levels l_1 and l_2. The different functions f_1 to f_3 represent three different combination methods.

Restrictions to the tree include the number of input forecasts as well as limitations in the number of steps, the maximal number of forecasts to combine per combination procedure and the set of potential combination methods.

3.1.3 *Definition of the optimum criterium*

The most simple and intuitive criterium to optimize is defined by the quality of the resulting forecasts. We want to learn combination structures which generate high quality combined predictions measured on an previously defined and unseen testbed (details will be given in the experimental section). The error is calculated as a mean absolute deviation value on the level of forecasting and is given as

$$E = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} |\widehat{x}(t) - x(t)|, \tag{2}$$

where t represents a time index over the evaluation period $[t_1, t_2]$. More sophisticated versions can include additional information like errors calculated at different levels or penalty terms corresponding to the complexity of the structure or the independence of the included combination procedures.

3.2 USING GENETIC ALGORITHMS

We started with the most common and simple approaches which are genetic algorithms. However it became clear quite quickly that a fixed length bit- representation of the objects to evolve are not ideal in order to represent our tree based dynamic structures. Even if the number of input forecasts to the combination process is restricted, we could not avoid to get chromosomes with a complex structure of genes if the size of potential steps is bigger than two and more than one combination model may be used.

3.3 USING GENETIC PROGRAMMING

A more flexible representation which is perfectly fitting to the tree based multi step combination structures is offered by the approach of genetic programming (see e.g. [9] or [10]). A genetic program (GP) can be interpreted as a tree with ordered branches, in which each node represents the application of a primitive function on arguments passed to the node by the branches from the next lower level. The leaves represent basic arguments called terminals. The root node represents the application of the function generating the final result.

The process of the development to evolve combination structures using GPs includes the following steps (see [10]):

1. Determine the set of terminals and select the set of primitive functions.
2. Define a fitness function.
3. Define an initial population.
4. Define crossover and mutation operators.

The next subsections follow these steps.

3.3.1 *Determine the set of terminals and select the set of primitive functions*

The terminals correspond to our chosen subset of the set of potential input forecasts $\{\widehat{x_l^m}\}$.

The set of primitive functions corresponds to the set $\{f_k\}$ of different combination methods included into the evolution process. If we want to use only one predefined combination model, we have only one primitive function describing a combination procedure.

3.3.2 *Define a fitness function*

The fitness function is modelled as a direct representation of the criterium to optimize, which means that it represents the quality of the resulting forecasts in terms of a mean absolute forecast deviation value and potentially including other information and/or penalty terms. Note that this definition makes the evaluation of the fitness function expensive in terms of performance, because it may contain a new calculation of the weights or parameters of all combination procedures included in a combination structure as well as a determination of error and correlation terms on the testbed.

3.3.3 *Define an initial population*

The population size is limited because of computational power and performance. As each member of the population represents a combination structure consisting of different combination procedures for which the combination weights (or other parameters if we have a nonlinear combination model) have to be learned for fitness evaluation, the population size should be as small as possible in order to be able to run the evolution quickly. On the other hand we have to assert that the space of potential solutions is well covered, at least in the domain where we can expect the optimal solution. That is the reason why it can be worth focusing on the determination of good initial populations.

We have followed two strategies which were both based on input forecast selections as described in 3.1 . In the first strategy we generated initial combination structures randomly only based on a few parameters like e.g. mean value and standard deviation given for the number of input forecasts for each combination procedure, the number of steps or the number of combination procedures to include per step.

In the second strategy we used our knowledge gained during the process of constructing the intuitive structures described in section 2.2. We used different of our intuitively predefined structures as initial populations for the evolution of the dynamic structures. The initial fitness following this second approach was slightly better than the one we achieved on average following the first approach, but we could not achieve significantly better results after the evolution.

3.3.4 *Define crossover and mutation operators*

We can here use the standard operators described e.g. in [10]. The crossover operator randomly exchanges subtrees of the two parents, for our combination structures meaning that there we exchange substructures or single combination procedures. For our problem the crossover operator has to be restricted in the sense that limitations of the maximal number of steps are not violated. Very stable versions of crossover allow only exchanges of subtrees representing the same step of combination.

The mutation operator randomly exchanges a terminal or a primitive function. Concerning the combination structures, mutation means that the combination methods are changed in the combination procedures or that input forecasts are randomly exchanged in the combination procedures of step 1 (including the possibility to add or to remove an input forecast).

4 EXPERIMENTS

4.1 DESCRIPTION OF THE EXPERIMENTS

4.1.1 *Description of the testbed*

The experiments have been carried out with the objective to apply different combination techniques to seasonal demand forecasts of a German air carrier in order to improve the forecast quality.

The chosen testbed included **10 representative origin- destination pairs (ODs)** containing 2 transatlantic flights to America, 4 intercontinental flights to Asia, 2 European flights and 2 national flights as well as direct routings and routings containing more than one segment.

The experiments started with historical data of 2001. Later the investigations have been extended to data containing 2 time periods : the period from 1 of January 2001 to 1 of March 2002 and from 1 of October 2002 to 1 of April 2004. The data of the period between March 2002 to September 2002 has not been available.
The data included

- the number of bookings of the testbed on the fine level ODIFPOS (routing, combination of flights on that routing, departure date, fareclass, point of sale) at different times prior to departure,
- the availability information of the testbed on the fine level,
- an historically measured seasonal behavior at the departure per calendar week on a fixed level and
- expert estimations concerning market changes.

4.1.2 *Description of the selected input forecasts and combination methods*

All the reported forecasts are 8 weeks ahead out-of-sample forecasts.

Our first experiments have been carried out on not decomposed data (see also [7] and [11]). In order to illustrate the importance of adaptation in the discussed application firstly seven different, well known time series forecasting techniques have been applied to the above described data and their performance compared. Both the methods and their performance on the testing data are shown in Table 1. Mean absolute error per departure shown in the second column of the table have been used as the performance measure. As the mean demand at the departure for the examined data is about 2.6 bookings, the forecast errors vary between about 80% to 120% of the demand. As can be seen from Table 1 the best results have been obtained using the simple exponential smoothing method.

A highly specialised adaptation to seasonal behavior and other changes as well as to the current booking values is already existing at the current forecasting system (for experiments related to different adaptations see [11]). This adaptation leads to a forecast improvement of about 11%. In order to measure the impact of different combination approaches on an overall forecast quality, in the following experiments the quality of certain approaches is always given in terms of a percentage forecast improvement compared to this best performing individual forecasting method (i.e. the simple exponential smoothing method together with the adaptation). The performance has been measured on unseen data meaning a time period which did not correspond to the period of history building for the forecasts, to the period used to calculate combination weights of the combination procedures or to the period used for fitness evaluation of the dynamic structures.

As the use of combination models on the toal demand predictions has not been very successful (only 1% of improvement could be achieved), the following experiments were focused on the use of combination models combining forecasts calculated on decomposed data.
Five different combining models have been used (for details see [2] and [12]):

Forecasting method	**Mean absolute error**
Simple exponential smoothing	2.05
Brown model (without season)	2.13
Holt winters model (without season)	2.19
Linear regression	2.31
Autoregression	2.48
Moving average	2.53
ARMA	3.15

Table 1: Mean absolute error for different time series forecasting techniques without adaptation.

- (av) : the simple average model
- (outp) : the outperformance model as an example of rank based models
- (opt) : the optimal model with assumption of independence as variance based model
- (sigm) : a nonlinear functional approach expecting that different methods work well for low seasons and different for high seasons. The parameters of the sigmoid functions used are learned using evolutionary strategies.
- (anfis) : the nonlinear neuro-fuzzy approach ANFIS

The special focus of our experiments was put on the component predicting seasonal behaviour. The generated predictions differed concerning the sources of information, the functional approach to predict the future behavior, smoothing parametrisation for the preprocessing and the level of decomposition meaning the level on which the calculation of the current seasonal factor was carried out.

We included two sources of information: the current season based on booking information already existing for the flights departing in the future and the historical season measured based on the behavior of the previous years. Three functional approaches m_1 to m_3 were included into the experiments, m_1 representing an additive approach, m_2 an multiplicative approach and m_3 a no season assumption.[1] The decomposition was carried out on different levels concerning aggregation over day of week (DOW), compartment(COMP), fareclass(F) and point of sale(POS).

We have chosen a fix subset of all potential input forecasts covering well the whole space. Table 2 shows a list of the 12 included input forecasts.

Table 2: Input forecasts for seasonal behavior included into the evolution of dynamic structures.

nbr	source of inform.	funct. approach	parametrisation	decomp.*
0	current season	m_1	no smoothing	-
1	current season	m_2	slightly smoothed	-
2	historical season	m_1	strongly smoothed	COMP
3	current season	m_2	slightly smoothed	DOW
4	current season	m_2	slightly smoothed	POS
5	current season	m_2	slightly smoothed	F
6	current season	m_2	slightly smoothed	DOW, POS
7	current season	m_2	no smoothing	DOW, F, POS
8	current season	m_2	no smoothing	DOW
9	current season	m_2	strongly smoothed	-
10	current season	m_1	strongly smoothed	-
11	-	m_3	-	-

(*Listed are the dimensions which have been aggregated for decomposition.)

[1]Please note that due to the commercial sensitivity of the covered material both the data and the methods covered in this section could not be described in any greater detail.

Table 3: Experimental results using different structural approaches and different combination models. The first column indicates the structure of the combination process. The second column indicates if multilevel forecasts have been included or not. Then the results are represented as an improvement percentage for the different combination methods. The last column indicates that the structure included the choice of any linear combination method.

structure	ml	av	outp	opt	sigm	anfis	flexible
flat (8 input fc)	no	-5	-2	-3	-15	-11	-
flat (best subset of 5 input fc)	no	-4	-1	+1	-3	-2	-
flat (12 input fc)	yes	-3	+1	-2	-1	-10	-
flat (best subset of 5 input fc)	yes	-3	+3	+2	+5	+2	-
static (best found structure)	yes	-3	+5	+4	+9	+6	-
dynamic (2 steps, restricted crossover)	yes	+6	+11	+9	-	+2	+10
dynamic (up to 4 steps, restricted crossover)	yes	+6	+9	+9	-	-2	+7
dynamic (up to 4 steps, flexible crossover)	yes	+5	+5	+4	-	-15	-1

4.1.3 *Description of the parameters for controlling the run*

For the dynamic approaches we have chosen a population size of 20 chromosomes. The maximal number of forecasts to combine has been set to 5 as well as the number of combination procedures per step. The maximal number of steps varied between 2 and 4. The maximal number of iterations has been set to 100. The learning process is also stopped if the fitness has not been changed over the last 20 generations. We have learned structures using a single combination method as well as structures detecting the combination methods automatically out of the three included linear combination methods (in the experimental results the method is referred as ”flexible”).

4.2 EXPERIMENTAL RESULTS

We have carried out a number of experiments to compare combinations based on the different static and dynamic structures. Table 3 shows the resulting percentage improvement.

4.3 ANALYSIS OF THE RESULTS

It can be seen that for the flat structures the best results can be achieved using nonlinear combination models on multilevel forecasts, but that there are also more relevant losses based on a higher risk of instabilities. The results achieved using these structures also show that a reduction of the chosen input forecasts to a subset of up to 5 forecasts leads to much better results, especially for the more sensible combination methods like the nonlinear methods.

The dynamic structures clearly outperformed the static structures, the results achieved using the simple average model or the outperformance model are surprisingly good. But the fact that the best result have been achieved using the most restricted version of crossover and using the most stable combination models also indicate that we still have a problem of overfitting in the more flexible cases. The same reason can be responsible for the relatively poor results of the structures which included the automatic choice of the combination method.

This effect has been verified by a very simple analysis of the achieved percentage improvements compared to the number of steps or the number of combination procedures. It clearly showed that the bigger structures achieved poor results because of missing generalisation capabilities. These structures were often learned in situations were the number of bookings was very low on the fine level which indicates that an extension of the fitness function with error terms measured at higher levels could be advantageous.

It was also very interesting to see that the stable structures seemed to have a tendency to generate structures which clustered the input predictions depending on their kind of diversity similar to the intuitive approach used when building the static structures. So we could observe a clear tendency to combine first different forecasts generated at the same level but using different functional approaches and then to combine the forecasts representing the different levels or visa versa.

Exceptions could often be explained by stability reasons. Especially in cases of very small numbers we could often achieve structures clustering more stable forecasts of lower levels e.g. achieved with a stronger smoothing during preprocessing, with more flexible forecasts coming from higher levels.

5 CONCLUSIONS AND FUTURE WORK

We have seen that dynamic multilevel combination models learned with the approach of genetic programming can be a powerful approach in order to build a high quality and adaptive forecast system. We have compared models differing concerning decomposition and multilevel combination structures and have achieved an improvement of forecast quality up to 11 percent. The best results have been achieved with highly restricted dynamic multilevel structures based on combination procedures using very simple and stable linear combination methods.

Further investigations are related to the generation of more sophisticated and specialised versions of fitness calculation and the crossover and mutation operators. We also continue the research in the direction of the choice of appropriate initial populations as well as an automatic adaptation of the pool of included input forecasts during the evolution.

References

[1] Russell, T.D. and Adam,E.E., 1987, ”An empirical evaluation of alternative forecast combinations”, European Journal of Operations Research 33, 1267-1276

[2] De Menezes, Lilian M. et al., 2000, ”Review of guidelines for the use of combined forecasts” Management Science 120, 190–204

[3] Shan, M.S. et al., 1999, ”Improving the accuracy of nonlinear combined forecasting using neural networks”, Expert Systems with Applications 16, 49–54

[4] Fiordaliso, A., 1998, ”A nonlinear forecasts combination method based on Takagi- Sugeno fuzzy systems”, International Journal of Forecasting 14, 367–379

[5] Bunn, E.W., 1985, ”Statistical efficiency on the linear combination of forecasts”, International Journal of Forecasting 1, 151–163

[6] Merz, C.J. and Pazzani, M.J., 1997, ”A Principal Components Approach to Combining Regression Estimates”, Machine Learning 0, 1–25

[7] Riedel, S. and Gabrys, B., 2004, ”Hierarchical Multilevel Approaches of Forecast Combination”, Proceedings of the GOR 2004 conference, 1-3 September 2004, Tilburg, Netherlands.

[8] Winkler,R.L. and Makridakis,S., 1983, ”The combination of forecasts”, Journal of the Royal Statistical Society 146,150–157

[9] Koza, J.R., 1992, ”Genetic Programming: On the programming of computers by means of natural selection”, MIT Press, Cambridge, MA.

[10] Negnevitsky, M., 2005, ”Artificial Intelligence: A guide to intelligent systems”, Second edition, Addison Wesley, ISBN 0-321-20466-2.

[11] Riedel,S. and Gabrys, B., 2003, ”Adaptive Mechanisms in an Airline Ticket Demand Forecasting System”, Proceedings of Eunite 2003, Oulu, Finland, 10-12 July 2003.

[12] Jang, J.-S.R., 1993, ”ANFIS: Adaptive Network- based Fuzzy Inference Systems”, IEEE Transactions on Systems, Man and Cybernetics 23, 665–685