Composite Forecasting Strategy Using Seasonal Schemata

Lawrence Beadle and Colin G Johnson
Computing Laboratory, University of Kent, Canterbury, CT2 7NF

***Abstract* - The essence of forecasting is to model patterns within a given set of data and reproduce these patterns in order to create a more accurate forecast. Different data sets will present different kinds of patterns and whilst many different methods and hybrid methods have been developed to model these different patterns, there have been attempts with some success to develop a method which will combine these methods and evaluate their strongest results as components and bring them together to create a forecast.**

This paper describes a new method for forecasting (*a composite forecasting strategy using seasonal schemata*) that learns a seasonal schema made up of different forecasting methods and uses this schema to create a combined forecast. This technique does not use the weights system of combination as first proposed in Bates & Granger's paper in 1969. Instead, we describe and evaluate a novel technique for combining different forecasting methods and demonstrate the results of this technique against a traditional single best forecasting method.

I. Introduction

This paper describes a novel method of forecasting, which works by decomposing the results of individual forecasts by seasonal segment and using the errors for each season to create a seasonal schema of the best results from the best forecasting methods and relate these methods to the relative seasonal period. In theory this should combine the best results for each season and enable the composite forecasting strategy to be more effective than standalone techniques.

Different forecasting techniques are better at dealing with different data characteristics; for example, some methods are better at dealing with linear trends compared to seasonal variations, whilst other methods may be better at dealing with seasonal factors rather than trends. An experienced forecaster will be able to apply their knowledge to make a decision about which method to apply in which situation. In the composite forecasting strategy uses a machine learning method to automatically discover a schema of which method to apply at each point in a cycle.

Traditionally, a number of techniques are used to create a forecast and the single method that produces the least error during the period where the forecasted data can be compared to the original data is chosen to produce further forecasts. This paper attempts to take this further by evaluating which methods are best for the individual seasons and basing its forecasts on this seasonal selection.

The contribution that this paper and software makes is the novel technique (*composite forecasting strategy using seasonal schemata*) used to combine the different forecasting methods. We seek to decompose error results by season, evaluate these errors and develop a forecast using a schema which will define which forecasting method should be applied in which season of the forecast.

The remainder of this paper contains a brief review of existing forecasting techniques, a full description of the workings of the composite forecasting strategy using seasonal schema with a small worked example on the combination method and a full worked example with results, discussion and evaluation.

II. Existing Forecasting Techniques

There are a variety of existing forecasting techniques. In the first instance there are the time series techniques such as naïve methods, decomposition, exponential smoothing, multiple regression and the ARIMA Box-Jenkins methodology. All of these methods are well established and detailed descriptions can be reviewed in Makridakis, Wheelwright and Hyndman's book [3] and Hanke, Wichern and Reitsch's book [4].

The combination of different forecasting methods was first proposed by Bates and Granger in [1], used extensively by Winkler and Makridakis in [2] and further discussed by Hendry and Clements in [5]. This combination method uses a system of weights applied to each method to create an overall forecast value. Whilst the method we propose does have the overall objective of combining forecasting methods, it works in completely different way to methods described in the above papers.

It is worth knowing that in Bates and Granger's 1969 paper [1] one of the conditions imposed is "…we impose one condition on the nature of the individual forecasts, namely that they are unbiased.". The method we propose should not be affected by bias through combining forecasting methods as it has the ability to discard methods on a season by season basis. This will result in any bias being minimized for each season.

In other work proposed by Harrison [6], based on an idea with similar properties to that of a Kalman filter [7], a method which combined auto regressive models was described. The method we propose does not use the same "measurement update" part of the time update, measurement update Kalman filter cycle.

There are several other techniques which have not been added into the current system and used in this investigation. These include evolutionary algorithms such as neural networks [8], [9] and genetic programming [10]. There have been some attempts to create hybrid models [11], however, at the time of writing, there does not appear to be a similar system to the one proposed in this investigation.

There are other methods that have been used to create forecasts such as collective intelligence approaches [12] and using textual web data to develop forecasts [13]. These techniques are standalone techniques which do not use composite strategies as proposed in this paper.

III. Developing a Composite Forecasting Strategy Using Seasonal Schema

This section describes the new composite forecasting strategy. Firstly, methods for extracting the seasonal schema from the data are described. Then the seasonal schema is applied to calculate the overall forecasts for each different seasonal point. Finally, methods for evaluating the algorithm are discussed.

A. Establishing a Seasonal Schema

In this investigation a software system has been implemented based on eight time series techniques. These were a naïve model, double and Holt-Winter's exponential smoothing, additive and multiplicative decomposition, logarithm multiple regression, multiple regression and a segmented least squares method which pulls the data apart by season, analyses the trend and then recombines the seasonal forecast. The system has been implemented in such a way that more methods can be added and the system will be able to evaluate them dynamically.

The seasonal schema quite simply represents which method is best to use is which seasonal period, for example, in the first quarter and second quarters use multiple regression, for the third quarter multiplicative decomposition and for the fourth quarter, Holt Winter's exponential smoothing.

Our software system creates a table of absolute errors for each method used for each piece of data in the fitting period. Once this is complete, the system calculates a total of absolute errors for each season for each method. Once this is complete, the system chooses one method that has the lowest absolute error for each season. The system then stores this information as a *seasonal schema*.

B. Applying the Seasonal Schema

Once the schema has been established, the software then creates the overall forecast by pattern matching the fits and forecasts from the correct method to the correct seasonal period based on the seasonal schema. In order to illustrate this technique, the following section shows a worked example of the application of this pattern matching technique.

C. Software

A software tool that uses multiple forecasting techniques, selects the best single method, and creates the pattern matched forecast has been implemented. This tool will also produce the statistical measurements of all of the forecasts and the pattern matched results based on the seasonal schema.

D. Measuring Fitting Period Accuracy

In the NN3 competition the measure of SMAPE (Symmetric Mean Absolute Percentage Error) has been chosen to compare the different forecasting techniques. Results in this paper are published in terms of SMAPE for easy comparison.

IV. Worked Example – Obtaining Seasonal Schema

This small worked example represents a data set of twelve values of quarterly data.

Table of Absolute Errors

Time Period	Season	Method 1 Error	Method 2 Error	Method 3 Error
1	1	1	2	3
2	2	2	1	4
3	3	3	3	1
4	4	4	4	1
5	1	1	2	4
6	2	4	1	3
7	3	3	3	1
8	4	2	4	1
9	1	1	2	4
10	2	2	1	3
11	3	3	3	1
12	4	4	4	1

Table 1

Table one shows example absolute errors for each one of three example methods over each time period during data fitting.

Absolute Error Totals

Season	Method 1 Total Error	Method 2 Total Error	Method 3 Total Error
1	3	6	11
2	8	3	10
3	9	9	3
4	10	12	3

Table 2

Table two shows the total absolute errors for each method for each seasonal period. The resulting seasonal schema would be:

RESULTING SCHEMA

Season	Method Used
1	Method 1
2	Method 2
3	Method 3
4	Method 3

Table 3

Table three shows the resulting seasonal schema. The software will then use this schema to select the fitting and forecast values used to create the composite forecast.

V. WORKED EXAMPLE – NN3_101 DATA SERIES

This example is based on the NN3_101 data series in the NN3 competition submission. There were 126 records and it is not know what this data represents. We are told that the data has 12 seasonal periods.

SMAPE OVER FITTING PERIOD

Method	SMAPE
Holt Winter's Exponential Smoothing (Single Best Method)	2.099371
Composite Method Using Seasonal Schema	2.081136

Table 4

Table 4 shows that the single best method for forecasting the NN3_101 data series was the Holt Winter's Exponential Model. This produced a SMAPE over the fitting period of 2.099. The composite method was able to reduce this slightly to a SMAPE of 2.081. It is worth mentioning at this low level of error it is relatively difficult to make a big improvement. The methods that were used to create the forecast which was pattern matched to the season schema are shown in table 5.

SCHEMA FOR NN3_101

Season	Method Used
1	DecompositionMultiplicative
2	HWeS
3	HWeS
4	HWeS
5	HWeS
6	HWeS
7	DecompositionAdditive
8	DecompositionMultiplicative
9	DecompositionMultiplicative
10	HWeS
11	HWeS
12	HWeS

Table 5

Table 5 shows that three different forecasting methods were used during the fitting period. These methods were Holt Winter's Exponential Smoothing, Multiplicative and Additive Decomposition. The schema also shows that the Holt Winter's method is used in eight out of twelve of the seasonal periods. In the four periods that the Holt Winter's model was not used it would imply that the Holt Winter's model had failed to take into account some element of pattern within the data, which another method was able to take into account.

Whilst this example has demonstrated that the composite strategy can work, the improvement in the SMAPE is marginal. It would be important at this point to examine one more case study where the single best method may not be as accurate as the NN3_101 data set result in order to see if better results can be produced in the face of greater errors in the fitting period.

VI. SECOND WORKED EXAMPLE – NN3_030

The NN3_030 data set proved slightly harder to model for the single best method which was multiplicative decomposition.

SMAPE OVER FITTING PERIOD

Method	SMAPE
Multiplicative Decomposition (Single Best Method)	5.524711
Composite Method Using Seasonal Schema	4.446832

Table 6

Table 6 shows over a 1% improvement in SMAPE, which when compared to a 5.5% SMAPE (for the single best method) is a noticeable improvement in the ability to model pattern within the software. In this case the methods used were also more diverse.

SCHEMA FOR NN3_030

Season	Method Used
1	MultiRegression
2	DecompositionMultiplicative
3	MultiRegression
4	MultiRegression
5	DecompositionAdditive
6	HWeS
7	Naive
8	HWeS
9	HWeS
10	LogMultiReg
11	MultiRegression
12	DecompositionMultiplicative

Table 7

Table 7 shows that whilst the decomposition method does provide the best result it only features twice in the schema. This is important, because, in this case, it is apparent that other models are performing well in terms of modeling the pattern of the data set in the fitting period. This implies that there is a benefit from combining the different forecasting models in this way.

VII. Analysis

The two examples in the previous section have shown that this technique can work effectively. At this stage instead of blindly printing the results of all 111 data sets from the NN3 competition, it would be more useful to consider some of the issues that could explain why this method works as well as how it could fail.

A. Why the Composite Method Works

This composite technique works by using multiple method to model more of the characteristics of the source data set. A simple graphical representation of the residuals from the fitting period of the NN3_030 data set shows an interesting feature.

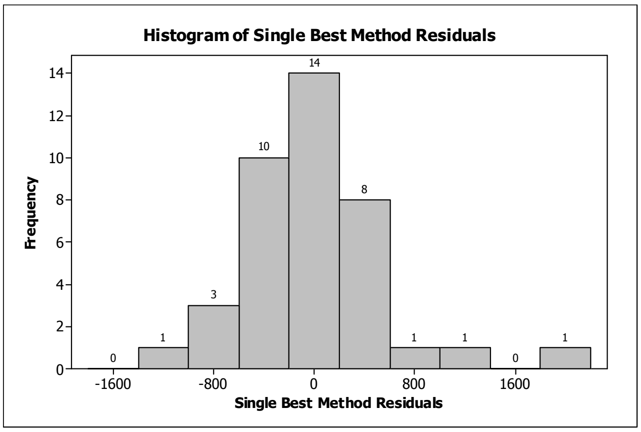

Figure 1

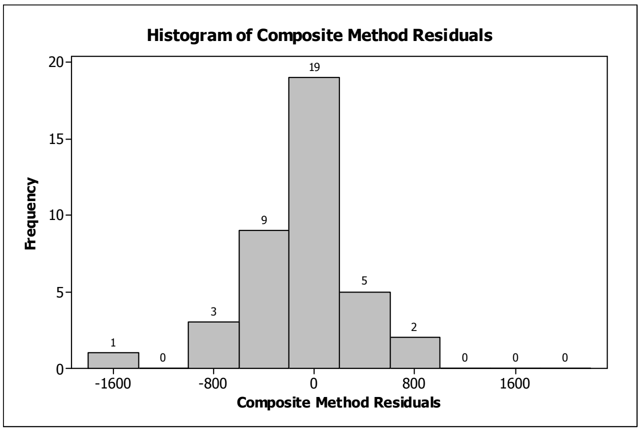

Figure 2

Figure 1 and 2 show that the residuals are essentially pushed more towards zero when the composite strategy takes effect. On examination of the relative difference between the two previous mentioned cases, one could assert that, if more methods are used then there will be more compression in the range of the residuals.

B. Is there a Case for Failure?

Previous experimentation with the software revealed that in the worst case scenario, the schema would simply revert to comprise solely of the single best forecasting method. In this situation it could be argued that it is not the schema creation ability that has failed, as it has still created the best performing schema possible. At this point, the failure becomes a reflection of the limitation of all the forecasting models in the system and their ability to model the data characteristics within the source data set. It could also imply that only one model had any ability to model the data characteristics presented in the source data set.

VIII. Discussion

There are two other factors that could affect the performance of the pattern matching software besides those already discussed.

The first of these is the season length. It is noted from previous experiments with quarterly data that the pattern matching from the seasonal schema seems to work better the larger the season length (i.e. increasing the value of s).

We could make an approximation of the probability of using one method in a seasonal period, if we make the assumption that, all forecasting models within the system are equally useful. If there are more seasons, it is more likely that one method will not out perform every other all of the time. It could be said that the probability of using one method all the time could be expressed as:

$$P = \left(\frac{1}{n}\right)^{s}$$

Where:

P = Probability of using one method for the whole forecast

n = The number of methods (in this case 8)

s = The number of seasons (in this case 12)

This would result in a very small probability of only using one method for an entire schema. In reality, however, it does not work in this way because if one method performs well then there is a better chance of it performing well (i.e. it is not mutually exclusive) in another season so this would have to be taken into account. All can be deduced from the above expression is that the bigger the s and n values the smaller the probability of using one method for all seasons, therefore, the better the probability that the composite strategy will work.

This simple expression would also suggest that, if the number of forecasting methods within the system is increased, the probability of using the same method will decrease. It has already been discussed that the worst case schema is one forecasting method, therefore the schema analysis software could be more successful each time a new method is added.

The second factor that could affect performance is the number of records in the actual data. The more data the statistical models have to analyse the patterns of, the more chance there is of recognising a more accurate pattern. This means that the more records there are, the models will have more chance to evaluate the pattern and as a result there should be two or more models with close results, which would mean that the pattern matching improvement may not be that much of an improvement. This area would be extremely difficult to conduct a full investigation, because it depends on the data sets used. It would require a large data set to retain consistent characteristics throughout and avoid any erratic fluctuations.

IX. CONCLUSIONS

In conclusion, it has been shown that the composite forecasting technique can reduce error in the fitting period and therefore should be more accurate in the forecast period. Further evaluation has shown that the composite technique using the seasonal schema could analyse more of the pattern within the data and as a result produces fitted predictions with less error compared to the fitted predictions of the single best forecasting method.

The composite strategy does not fail in the sense that it performs worse than the single best method because even in situations where only one method is used it still picks that model and the results of that model to construct the pattern matched forecast based on the seasonal schema.

The reason the pattern matching software fails to create a pattern other than the single best method is that the second best performing method is relatively inaccurate compared to the best single forecasting method. This means that the probability of success in terms of the composite strategy producing a lower SMAPE is dependant on two or more forecasting methods being comparable in terms of fitting period error.

Whether or not two methods are comparable in terms of error depends on the actual data being forecasted. As discussed, this would mean that conducting a hundred tests and giving a percentage result of the number of test cases that achieved a better SMAPE using the composite strategy. This may not be a valuable statistic because with one hundred different sets of data the percentage of cases attaining an improvement using the composite strategy could be totally different.

It has also been discussed how the season length, number of forecasting methods being used and the number of records may affect the performance of the composite forecasting software. The greater the season length and number of methods used, the greater is the chance that the schema creation software will be able to recognise a schema of methods to use other that the single best method.

Finally, the one thing that can be taken from this investigation is that, if the worst the schema creation software can do is give the results of the single best method and the best is reduce the fitting period error, therefore reducing the forecasting error, then it should always be used when forecasting because it is not possible to do any worse than the single best method result.

X. RECOMMENDATIONS FOR FURTHER WORK

A first area of consideration is to add more methods to the system. The system dynamically uses the forecasting methods (through Java.reflect) as long as they are structured to a set specification. Implementations of the ARIMA models and Neural Nets would benefit the system. Also, there are other ways of implementing methods that are already present. Note that, apart from the computational time cost, there is no disadvantage to adding new methods to the system.

Another interesting study would be to allow irregular factors to be entered to help analyse patterns in the data. A good example of this would be profiling of promotional offers in industry, perhaps using marketing models, so it would be possible to enter the start time period and the expected magnitude of the promotion and as a result account for promotional patterns or other irregular patterns in fits and forecasts.

It would be interesting to redesign the schema analysis part of the software so it would look for alternate methods (aside from analysis absolute errors) of creating the seasonal schema. One such example would be to develop a system of penalising larger residuals first and then working to decrease the number of smaller residuals.

LIST OF REFERENCES

[1] J. M. Bates and C. W. J. Granger, "The Combination of Forecasts," *Operational Research Quarterly*, vol. 20, pp. 451-468, 1969.

[2] R. L. Winkler and S. Makridakis, "The Combination of Forecasts," *Journal of the Royal Statistical Society*, vol. Vol. 146, pp. 150-157, 1983.

[3] S. G. Makridakis, *Forecasting : methods and applications*, 3rd ed / Spyros Makridakis, Steven C. Wheelwright and Rob J. Hyndman. Translator: Wheelwright, Steven C., Hyndman, Rob J. ed. New York ; Chichester: Wiley, 1998.

[4] J. Hanke, D. Wichern, and A. Reitsch, *Business Forecasting*, 7th Edition ed: Prentice Hall Inc, 2001.

[5] D. F. Hendry and M. P. Clements, "Pooling of forecasts," *The Econometrics Journal*, vol. 7, pp. 1-31, 2004.

[6] R. Harrison, G. Kapetanios, and T. Yates, "Forecasting with Measurement Errors in Dynamic Models," B. o. England, Ed., 2004.

[7] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. Vol 82, pp. 35-45, 1960.

[8] Z. Tang and P. Fishwick, "Feed-Forward Neural Nets as Models for Time Series Forecasting," 1993.

[9] T. F. a. M. Vornberger, "Sales Forecasting Using Neural Networks," *Proceedings ICNN'97*, vol. Volume 4, pp. 2125-2128, 1997.

[10] C. Grosan, A. Abraham, S. Han, and V. Ramos, "Stock Market Prediction Using Multi Expression Programming," presented at 12th Portuguese Conference on Artificial Intelligence (EPIA'05), Portugal, 2005.

[11] S. Singh and J. Fieldsend, "Pattern Matching and Neural Networks Based Hybrid Forecasting System," *Lecture Notes in Computer Science*, vol. 2013, pp. 72, 2001.

[12] C. A. Kaplan, "Collective Intelligence: A New Approach to Stock Price Forecasting," presented at IEEE Systems, Man, and Cybernetics Conference, 2001.

[13] B. Wuthrich, D. Permunetilleke, S. Leung, V. Cho, J. Zhang, and W. Lam, "Daily Prediction of Major Stock Indices from Textual WWW Data," presented at IEEE International Conference on Systems, Man, and Cybernetics, 1998.