Ensemble of Time Series forecasting in Complex Structure

Shanu Agrawal
Robert Bosch Engineering and Business Solutions Bangalore, India

Balasubramanyam Pisupati
Robert Bosch Engineering and Business Solutions Bangalore, India

Abstract

Forecasting is necessary for better business understanding and decision making. When it is done manually, it requires a lot of effort and time from multiple department like logistics, sales, finance etc. It also involves lot of gut feeling from experienced people and sometimes it might lead to error prone prediction if person is inexperienced or is not aware of past behavior. It is even very challenging for any data scientist to find one forecast model that performs best for all scenarios and in all forecast horizons. In this paper an approach for forecasting using ensemble model is discussed. Ensembling is done using Symmetric mean absolute percentage error (SMAPE) and mean absolute percentage error (MAPE) calculated from rolling forecast approach. For validation of forecast model, M3 competition data is used.

Keywords: forecasting, time series, ensemble, forecast combination

1. Motivation of Research

The decisions based on data plays crucial role in successful planning for any organization. Forecasting is one such widely used data based approach to decide the future scope of action. According to **Allende & Valle**'s paper: Ensemble techniques for time series forecasting have indispensable importance in many practical data mining applications. Combining forecasts from conceptually different methods is a very effective way to improve the overall precisions in forecast. It is an ongoing dynamic area of research, and over the years various forecasting models have been developed in the literature. It is not an easy task and so far no single model alone can provide best forecasting results for all kinds of time series data.

According to Oliveira's paper: The main motivation of this ongoing work is the observation that handling time series tasks requires several decisions in terms of how we describe the recent dynamics of the observed values of the series. Settling on a single answer to these decisions may be dangerous in real world time series where one frequently observes changes in the dynamic properties of the variable being measured. Ensembles are a well-known answer to this type of problems by taking advantage of diversity among models to reduce both the bias and variance components of the prediction error.

The objective of this paper is to automate forecast process using ensembling such that best accuracy can be achieved for all kind of scenarios where forecasting is possible and provide this as base model for planners to improve their decision making.

2. Literature Review

According to 'Forecast combinations in R using the forecast combinations package Manual' written by Eran: When it comes to forecasting, it makes model-selection uncomfortably restrictive in dynamic environments. And an obvious alternative to choose single best forecasting method is to combine forecasts from different models. In Oliveira's paper, similar statement has been made that ensembles are a well-known answer to this type of problems by taking advantage of diversity among models to reduce both the bias and variance components of the prediction error. By now it is well established that combining different forecasts will deliver result which is "better than the best".

In Spyros and Michel's paper it's proved that "The accuracy when various methods are combined outperforms, on average, the individual methods being combined and does very well in comparison to other methods". In this paper, various methods for error calculation is also discussed like SMAPE, Average Ranking, Median symmetric APE, Percentage Better, and Median RAE. They also proved that "The accuracy of the various methods depends upon the length of the forecasting horizon involved". It means performance of forecast model varies with forecast horizon, and different forecast model might perform well for different forecast horizon.

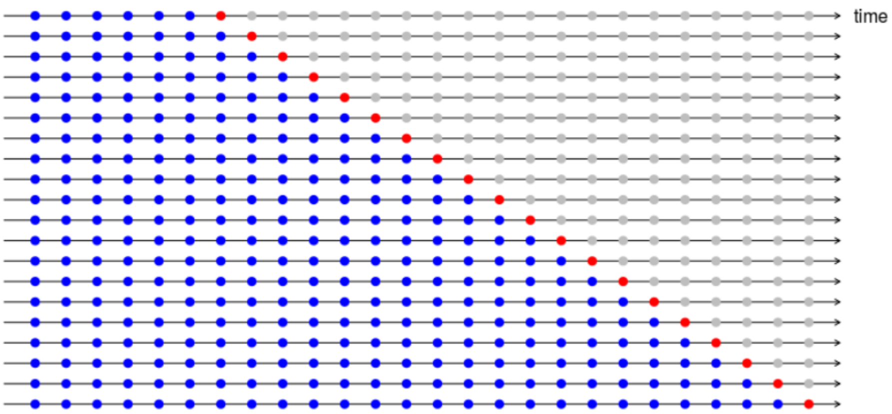
According to Allan's presentation at oxford university, even if combination of forecast do not always deliver the most precise forecasts, it generally do not deliver poor performance and so from a risk perspective represent a relatively safe choice. Due to this, forecast combinations outperform the best individual forecasting model.

In this paper, variance based combination and best Individual approach method with SMAPE and MAPE instead of mean square error used for ensembling. This approach assign more weight to more accurate forecasting models according to SMAPE/MAPE. Put simply, more accurate forecasting methods are weighted more heavily. Error is computed based on out of sample past forecasts, sometimes referred to as prediction Error.

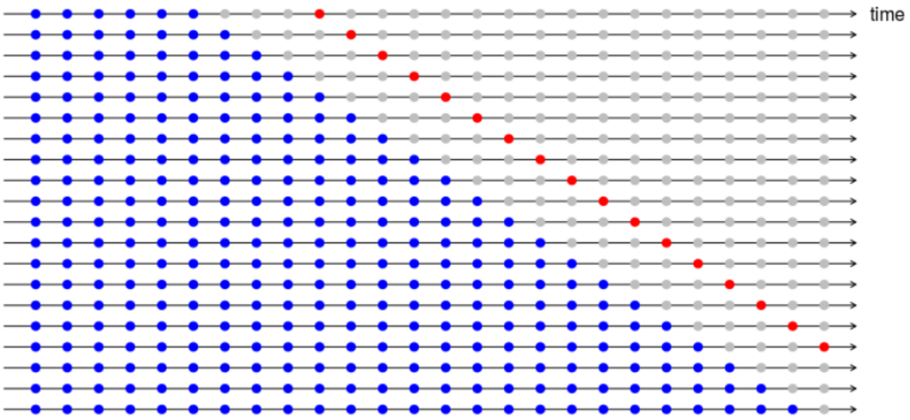
3. The Methodology

In this paper for forecasting, autoregressive integrated moving average model (ARIMA), exponential smoothing (ETS), naïve, snaïve, theta, and forecasting with decomposition (STL) are used as base models. After that multiple ensemble approach applied to combine the forecast from all models for better forecast on an overall.

Normally to check the efficiency of forecast model, the rolling forecast approach is used. It is a cross validation procedure for forecasting. In this procedure, there is a series of test sets, each consisting of a single observation. The corresponding training set consists only of observations that occurred *prior* to the observation that forms the test set. Thus, no future observations can be used in constructing the forecast. The following diagram illustrates the series of training and test sets. The blue observations form the training sets, and the red observations form the test sets.



The forecast accuracy is computed by averaging over the test sets. Suppose we are interested in models that produce good 4-step-ahead forecasts. Then the corresponding diagram is shown below. This approach is described by prof. Hyndman (https://robjhyndman.com/hyndsight/tscv/).



For each horizon, accuracy is computed by averaging over the test sets. In this paper, SMAPE and MAPE used for error calculation. Calculation for SMAPE and MAPE using the following equation:

$$\text{SMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{(A_t + F_t)/2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

where, n = number of test datasets, A_t = Actual value, F_t = Forecast value

The error calculation process is repeated for all mentioned forecast model. As proved Spyros and Michel's paper "The accuracy of the various methods depends upon the length of the forecasting horizon involved". It might be difficult for one particular forecast model to predict for all horizon accurately. The performance can vary with time because of dynamic behavior of time series data, it means one forecast cannot be best for all time. Some other forecast model might perform better than existing model, when behavior of time series varies. The best way is to combine forecast from multiple best models such that it performs for all forecast horizon and adopt accordingly for dynamic changes. This will automate the whole forecast process and help data scientist to find one forecast model that performs best for all scenarios and all forecast horizons.

In Eran's manual, combination models such as: Simple average, Ordinary Least Squares (OLS) regression, Least Absolute Deviation (LAD) regression, Constrained Least Squares (CLS) regression, Variance-based, or Inverse Mean Squared Error and best Individual model are discussed. In this study, combined forecast is calculated using the similar approach as inverse MAPE or variance based, instead of mean square error SMAPE and MAPE. This approach assign more weight to more accurate forecasting models according to error. The combined forecast is given by the following formula.

$$f^c = \frac{\frac{1}{\text{SMAPE}_i}}{\sum_{i=1}^{P}\frac{1}{\text{SMAPE}_i}} f_i \quad \text{Or} \quad \frac{\frac{1}{\text{MAPE}_i}}{\sum_{i=1}^{P}\frac{1}{\text{MAPE}_i}} f_i$$

Where p is the number of forecast models. For calculation, it requires past test forecast and actual value to derive the weight. Errors are computed based on past forecast errors of test data that are not used for training the model for red dots according to rolling forecast approach. This process will be repeated for each horizon. Once we get weight for each model, current training data forecast will be multiplied to get combined forecast as shown in figure 1. Similarly for horizon 2, using all previous training data of same horizon, weight will be derived for each model and then current period training data forecast will be multiplied to get combined forecast. This process will be repeated for all required horizon. To check the robustness of model, rolling approach can be used.

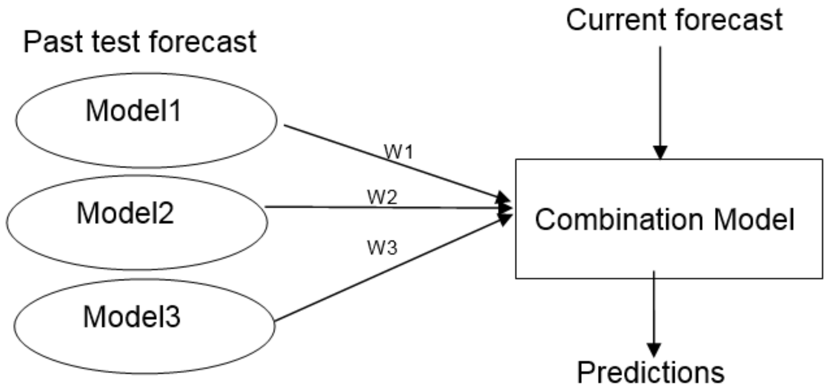


Figure 1: Approach for Forecast Combination Model

Using only SMAPE and MAPE for assigning weights work well when all forecast model have similar variation. But when forecast models have wide variance this might not work well, in such cases have to assign weight according to preciseness and reliability of model. Reliability of model is measured by variance. If any model have more error variance, it will be less reliable. The combined forecast is calculated by both mean error and standard deviation, using the following formula.

$$f^c = \frac{\frac{1}{\mathrm{MAPE}_i} + \frac{1}{s(\mathrm{APE}_i)}}{\sum_{i=1}^{P} \frac{1}{\mathrm{MAPE}_i} + \frac{1}{s(\mathrm{APE}_i)}}$$

Using the similar approach combined forecast is derived using SMAPE error and standard deviation. Instead of simple mean, weighted mean assigning more weight to recent forecast errors can be used if time series is more dynamic.
In many applications the time series organized in a hierarchical structure based on dimensions such as region, customer and product type. It is usually forecasted using grouped time series and requires forecast for each combination of series. It adjusts the forecast such that bottom up forecast matches with top down forecast. Behaviour of each series can be different from other and using one model for all series may not give good accuracy. In such cases combined forecast model works well.

4. The Findings

For study, M3 competition (https://forecasters.org/resources/time-series-data/m3-competition/) time series are used, category of data shown in table 1. Number of observation is different for each time series, it varies from 66 to 144 and also has different behaviour. The required forecast for each time series is 18 month ahead.

As mentioned in methodology, variance based approach is used with MSE, SMAPE, MAPE with and without standard deviation used to combine forecast from different models. To assign weight to each forecast model, errors are computed based on past six forecast of test data according to rolling forecast approach for monthly forecast. This process is repeated for each horizon. We created 25 rolling forecast, to get equal number of test forecast for all horizons. Once we get weight for each model, current training data forecast will be multiplied with weight to get combined forecast. Similarly combined forecast is calculated for Quarterly and yearly data of M3-Competition.

Table 1: Description of M3 Competition Data

Time interval between successive observations	Types of time series data						
	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57		756
Monthly	474	334	312	145	111	52	1428
Other	4			29		141	174
Total	828	519	731	308	413	204	3003

In table 2 average SMAPE for eighteen month ahead forecast for M3 competition monthly data is shown. We can observe that performance for combined models are good over the period of time. Combined models have minimum error on average than base models, thus we can draw an inference that the error is minimum for all horizons but not the least. However, since the model is applicable to forecast all horizons it is more convenient for data scientist to rely on a single model rather than opting for different model for different forecasting horizon. Similar result is observed for Quarterly and yearly data, in table 3 and 4, average SMAPE for 8 and 6 period ahead forecast for all categories is shown.

Table 2 Average SMAPE: M3 Competition Monthly Data All categories

Method	Forecast Horizon										#Obs.
	1	2	3	4	5	6	8	12	18	Avg. 1 to 18	
ARIMA	14.1	13.8	14.7	15.9	13.9	14.5	15.6	15.8	21.9	17.0	1428
Combined_MSE	14.0	12.5	13.8	14.9	12.4	13.2	14.0	14.7	19.5	15.4	1428
Combined_MAPE	14.0	12.5	14.0	15.1	12.4	13.1	13.9	14.7	19.3	15.4	1428
Combined_MAPE_SD	14.1	12.5	14.0	15.2	12.5	13.2	13.9	14.7	19.3	15.4	1428
Combined_SMAPE	14.0	12.5	13.9	15.1	12.3	13.1	13.9	14.7	19.3	15.4	1428
Combined_SMAPE_SD	14.0	12.5	13.9	15.1	12.4	13.2	14.0	14.7	19.3	15.3	1428
ETS	13.6	12.4	14.0	15.2	13.6	13.8	14.7	14.9	21.0	16.0	1428
NAIVE	19.0	17.8	19.5	22.2	19.5	18.0	19.3	17.6	24.2	20.5	1428
SNAIVE	18.7	16.4	17.6	18.5	17.5	18.1	17.0	17.6	22.3	19.0	1428
STL	14.7	13.5	14.9	16.1	14.5	14.9	15.4	15.7	21.6	16.9	1428
THETA	13.3	12.6	13.7	14.3	13.7	13.7	14.2	14.5	19.9	15.6	1428

Table 3 Average SMAPE: M3 Competition Yearly Data All Categories

Model	Forecast Horizon							#Obs.
	1	2	3	4	5	6	Avg. 1 to 6	
ARIMA	9.78	13.94	18.96	22.51	26.00	28.02	19.87	645
Combined_MSE	7.86	11.73	16.12	18.54	21.18	22.31	16.29	645
Combined_MAPE	8.01	11.80	16.15	18.38	21.23	22.31	16.31	645
Combined_MAPE_SD	8.22	12.01	16.31	18.51	21.46	22.28	16.47	645
Combined_SMAPE	8.00	11.75	16.09	18.35	21.19	22.25	16.27	645
Combined_SMAPE_SD	8.19	11.90	16.19	18.40	21.38	22.21	16.38	645
ETS	9.35	13.23	17.26	20.05	22.96	24.67	17.92	645
NAIVE	8.51	13.23	17.77	19.90	22.96	24.90	17.88	645
SNAIVE	8.51	13.23	17.77	19.90	22.96	24.90	17.88	645
THETA	8.71	12.56	16.86	18.81	21.45	23.34	16.96	645

Table 4 Average SMAPE: M3 Competition Quarterly Data All Categories

Model	Forecast Horizon									# Obs.
	1	2	3	4	5	6	7	8	Avg. 1 to 8	
ARIMA	5.3	7.1	8.2	9.9	10.9	12.1	12.7	13.9	10.0	756
Combined_MSE	4.6	6.4	7.2	8.7	9.5	11.0	11.5	12.4	8.9	756
Combined_MAPE	4.8	6.5	7.2	8.7	9.4	10.8	11.3	12.3	8.9	756
Combined_MAPE_SD	4.9	6.5	7.2	8.7	9.4	10.8	11.2	12.8	9.0	756
Combined_SMAPE	4.8	6.5	7.3	8.7	9.4	10.8	11.3	12.3	8.9	756
Combined_SMAPE_SD	4.8	6.6	7.3	8.8	9.4	10.8	11.2	12.8	9.0	756
ETS	5.1	6.9	7.7	9.4	10.2	11.6	12.4	13.5	9.6	756
NAIVE	8.2	9.2	10.2	9.2	11.9	13.5	14.5	13.7	11.3	756
SNAIVE	9.1	9.1	8.6	9.2	12.9	13.3	12.5	13.7	11.1	756
STL	5.1	6.7	7.9	9.3	10.6	12.2	13.3	14.1	9.9	756
THETA	5.3	7.0	7.8	9.0	9.7	11.1	11.6	12.4	9.2	756

5. Summary and Conclusions

Combine forecast ensemble approach using SMAPE and MAPE with and without standard deviation, is helpful to find one of the best forecast model that performs well for all scenarios and all forecast horizon. Combined forecast might not give most precise forecasts, but do not deliver poor performance and so from a risk perspective represent a relatively safe choice. Due to this, forecast combinations outperform the best individual forecasting model. Combined model quickly adopts to the new forecast model as and when it observes the changes in behaviour of time series. As the model finds best combination according to previous performance, the accuracy is high. Combined method reduces both the bias and variance components of the prediction error. This behaviour helps planners for preparing quick, informative and effective decisions and they do not need to depend on data scientist for updating the model.

References

[1] Eran Raviv, (2016) Forecast combinations in R using the ForecastCombinations package

[2] Spyros Makridakis, Miche`le Hibon, (2000) The M3-Competition: results, conclusions and implications, International Journal of Forecasting 16 451–476

[3] Rob J Hyndman, Y. Khandakar, (2008) Automatic time series forecasting: The forecast package for R. J. Stat. Softw. 27 (3), 1–22.

[4] V. Assimakopoulos, K. Nikolopoulos, (2000) The theta model: a decomposition approach to forecasting, International Journal of Forecasting 16 521–530

[5] N. Kourentzes, (2016) How to choose a forecast for your time series http://kourentzes.com/forecasting/2016/06/17/how-to-choose-a-forecast-for-your-time-series/

[6] Hyndman Rob J, Athanasopoulos George, Forecasting: principles and practice, https://www.otexts.org/book/fpp

[7] D. Barrow and N. Kourentzes, (2016) Distributions of forecasting errors of forecast combinations: Implications for inventory management", International Journal of Production Economics 177: 24-33

[8] Rob J Hyndman, George Athanasopoulos, and Han Lin Shang. (2015) hts: An r package for forecasting hierarchical or grouped time series

[9] Mariana Oliveira, Luis Torgo, (2014) Ensembles for Time Series Forecasting, JMLR: Workshop and Conference Proceedings 39:360- 370, 2

[10] Héctor Allende and Carlos Valle, (2016) Ensemble Methods for Time Series Forecasting, Springer International Publishing AG 2017

[11] Allan Timmermann, (2013) Forecast Combinations http://www.oxford-man.ox.ac.uk/ sites/default/ files/ events/combination_Sofie.pdf