Ensembles for Time Series Forecasting

Mariana Oliveira and Luís Torgo

LIAAD-INESC TEC / Faculdade de Ciências - Universidade do Porto
mrfo@inesctec.pt, ltorgo@dcc.fc.up.pt

Abstract. We describe a new type of ensembles that aims at improving the predictive performance of these approaches in time series forecasting. Previous theoretical studies of ensembles have shown that one of the key reasons for this performance is diversity among ensemble members. The key idea of the work we are presenting is to propose a new form of diversity generation that explores some specific properties of time series prediction tasks. Our hypothesis is that the resulting ensemble members will be better at addressing different dynamic regimes of time series data.

1 Our Proposal

Most existing approaches to time series forecasting use the most recent observed values of the series as predictors for the future values (usually known as an embed of the time series). These approaches require setting a key parameter - how many past values to include. Setting this parameter is not trivial most of the times as there may not exist *one* single correct answer. In effect, non-stationary series and the occurrence of different regime shifts along time may lead to the best value being clearly time-dependent. The key idea of our proposal[1] is that of using different sets of predictors (e.g. different embed sizes) within the members of an ensemble to inject some diversity that is related with specific properties of time series tasks. More specifically, given a maximum embed size k_{max}, in this paper we will consider:

- E a baseline standard bagging approach using the previous k_{max} values of the target variable as predictors
- E+S an extension of standard bagging by adding two extra predictors that try to convey extra information on the dynamics of the series, namely μ_Y and σ^2_Y, calculated using the values within the maximum embed
- DE an ensemble where one third of the models use the maximum embed, another third uses an embed of $k_{max}/2$ and the last third uses $k_{max}/4$.
- DE+S an ensemble similar to DE but all models will have have the μ_Y and σ^2_Y extra features, although calculated with the respective embed.
- DE±S a variant of DE+S where for each third, half of the models will use the extra statistics, whilst the other half will only use the respective embed.

[1] Further details and an implementation at `www.dcc.fc.up.pt/~ltorgo/DS2014`

2 Experimental Evaluation

The main goal of our experimental evaluation is to check whether the new variants of bagging are able to outperform standard bagging on time series forecasting tasks. Our baseline benchmark is standard bagging using the approach tagged as E in the list above. All five variants were compared using the same base data (the k_{max} past values) as training set, but some use it in a different way, e.g. by using only part of it or by using it to generate extra features.

All five alternative forms of bagging were tested on fourteen real world time series (details at the paper associated Web site). Mean Squared Error (MSE) was used as evaluation metric to compare the different approaches. In order to obtain reliable estimates of this metric we have used a Monte Carlo simulation. In our Monte Carlo experiments we have randomly selected 10 points in time within the available time intervals of each task. For each of these 10 random points we have used as training set the previous 50% observations and the following 25% cases as test set. All approaches were trained and tested using the same exact data to allow for paired comparisons. Wilcoxon signed rank tests were carried out to test the statistical significance (with $p-$value < 0.05) of the observed paired differences in MSE of the proposed approaches against the bagging baseline.

We have repeated our experimental comparisons using 4 different setups: (i) number of models in the ensemble (M); and (ii) value of the maximum embed used by the ensembles (k_{max}). Table 1 presents the overall results of the paired comparisons. The numbers in column "Wins/Losses" are the wins and losses of each variant against the baseline, on the fourteen problems. Between parentheses we have the number of statistically significant (95% confidence) differences.

M	k_{max}	Variant	Wins/Losses	M	k_{max}	Variant	Wins/Losses
1020	20	E+S	13 (11) / 1 (1)	1500	20	E+S	13 (10) / 1 (1)
		DE	7 (7) / 7 (3)			DE	8 (6) / 6 (3)
		DE+S	13 (10) / 1 (0)			DE+S	13 (10) / 1 (0)
		DE±S	14 (12) / 0 (0)			DE±S	14 (12) / 0 (0)
	30	E+S	11 (9) / 3 (2)		30	E+S	11 (9) / 3 (2)
		DE	10 (6) / 4 (3)			DE	9 (7) / 5 (3)
		DE+S	10 (5) / 4 (2)			DE+S	10 (7) / 4 (2)
		DE±S	10 (9) / 4 (2)			DE±S	10 (9) / 4 (2)

Table 1: Paired comparisons results.

These results clearly show a positive overall balance of our proposed method for adding time series-specific diversity to bagging. In particular, the DE±S variant achieves remarkable results when $k_{max} = 20$, as it always outperforms standard bagging. This is the variant that introduces more variability within the members of the ensemble, which somehow provides further evidence of the advantage of our proposal. Overall, these results are encouraging and provide clear indications of the added value of this research direction even though many more possibilities exist to increase the level of diversity.