Aggregation and ensembles: Principled combinations of data

Perhaps *the* major story in forecasting the 2012 election is the growing awareness of the benefits of aggregating multiple sources of data for improving prediction. Most prominently, polling analysts including Simon Jackman, Drew Linzer, and Nate Silver made the strong case, ultimately validated by the election results, that combining information from multiple polls will give a better picture of the electorate than any poll analyzed in isolation.

Our approach, ensemble Bayesian model averaging (EBMA), draws on this same basic intuition, even if the specifics differ considerably from the poll aggregation methods referenced above. Instead of seeking to create the single “best” forecasting model, EBMA aims to “combine the intuition, theories, and concepts implicit in all of the forecasting models presented in this symposium to make an accurate out-of-sample prediction” (Montgomery, Hollenbach & Ward 2012, p. 651).

While no single election outcome can validate the ensemble approach, the performance of EBMA in 2012 was encouraging. The final outcome, 51.3% of the vote for President Obama, was well within our 95% predictive credible interval [46.4%, 52.5%]. Our point prediction, the median of the predictive posterior, was 50.3%. Thus, our prediction was off by about 1%. Of the ten forecasts included in the ensemble, only the Abramowitz and Campbell (Trial-heat) models offered a more accurate prediction with absolute errors of 0.7% each. Indeed, we would like to especially give credit to these authors, whose accurate forecasts of the 2012 outcome contributed so much to the ensemble’s performance. The Trial-Heat, and especially the Abramowitz model were among the most heavily weighted in the ensemble. EBMA’s reliance on them was based in part on their previous accuracy, which they repeated this cycle.

EBMAs strong performance provides additional support for the notion that more information is better. Yet, it is important to note that EBMA does not simply aggregate forecasts without respect to their past performance. For the purposes of this symposium, we weighted forecasts based on the accuracy of each model’s pseudo-forecasts (their in-sample predictive ability). Elsewhere, we took the more stringent approach of evaluating forecasting teams based on their true out-of-sample forecast made in advance of elections (Montgomery, Hollenbach & Ward 2012a). In either case, models that provided less accurate forecasts received less weight. This discrimination is why EBMA in general outperforms more naïve approaches to aggregating forecasts. The mean prediction of the models included in our ensemble was 50.0 and the median was 49.4; the EBMA aggregation performed considerably better than either of these. At the same time it preserves information about the uncertainty of the ensemble average.

It seems easy to predict that ensemble averaging will continue to be a part of election forecasting, given Drew Linzer and Simon Jackman’s success this year, as well as the accuracy and popularity of Nate Silver’s efforts. Doubtless it will be advantageous – and fruitful – to continue to broaden the range of models included in the ensemble in the next symposium.

# Bibliography

Montgomery, JM, Hollenbach, FM & Ward, MD 2012, 'Ensemble Predictions of the 2012 US Presidential Election', *PS: Political Science & Politics*, vol 45, no. 4.

Montgomery, JM, Hollenbach, FM & Ward, MD 2012a, 'Say Yes to the Guess: Tailoring Elegant Ensembles on a Tight (Data) Budget', *Annual Meeting of the American Political Science Association*, New Orleans, LA.