Say yes to the guess: Ensemble forecasting with sparse data *

Jacob M. Montgomery
Department of Political Science
Washington University in St. Louis
Campus Box 1063, One Brookings Drive
St. Louis, MO, USA, 63130-4899

Florian M. Hollenbach
Department of Political Science
Duke University
Perkins Hall 326 Box 90204
Durham, NC, USA, 27707-4330

Michael D. Ward
Department of Political Science
Duke University
Perkins Hall 326 Box 90204
Durham, NC, USA, 27707-4330
corresponding author: michael.d.ward@duke.edu

August 7, 2012

*This work was supported by the Information Processing Technology Office of the Defense Advanced Research Projects Agency through a holding grant is to the Lockheed Martin Corporation [FA8650-07-C-7749].

Abstract

Bla bla blub

1 Introduction

Although accurate prediction of future events has not . EBMA mixture models seek to collate the good parts of existing forecasting models while avoiding over-fitting to past observations or over-estimating our certainty about the future. The hope is for greater accuracy as both the knowledge and implied uncertainty of a variety of approaches are integrated into a combined predictive probability distribution.

Specifically, the amount and quality of data for calibrating ensembles is. However, Raftery et al. (2005) first proposed EBMA mixture models as

It is now common for people to make predictions

The first constraint is the paucity of the data. Relative to weather and other stuff, there just aren't a lot of observations, mostly stemming from the fact that for any given time period there is exactly one outcome.

A second is the large number of forecasting efforts. For example, the economic forecasting survey has like a billion experts.

A final issue is the inconsistency with which forecasts are issued. For any given time window, there are many missing foreacsts. Moreover, it is not random with larger degrees of misssingness existing further back in time.

As an example, Table 1 represents nearly the entire population of legitimate forecasting efforts. We see all three of these things going on at the same time.

Table 1: Pre-election forecasts of the percent of the two-party vote going to the incumbent party in U.S. Presidential elections

	F	A	C	H	LBRT	L	Hol	EW	Cuz
1992	55.7	46.3	49.7	48.9	47.3				
1996	49.5	57.0	55.5	53.5	53.3		57.2	55.6	
2000	50.8	53.2	52.8	54.8	55.4	60.3	60.3	55.2	
2004	57.5	53.7	52.8	53.2	49.9	57.6	55.8	52.9	51.1
2008	48.1	45.7	52.7	48.5	43.4	41.8	44.3	47.8	48.1

Forecasts were published prior to each election by **F**air, **A**bramowitz, **C**ampbell, **H**ibbs, **L**ewis-**B**eck and **R**ice (1992), Lewis-Beck and **T**ien (1996-2008), **L**ockerbie, **Hol**brook, **E**rikson and **W**lezien and **Cuz**an. Data taken from CITE SILVER POST HERE.

We adjust the basic EBMA model by better handling missing data,

2 Notation and EBMA model

Assume a quantity of interest to forecast, $\mathbf{y}^{t^*}$, in some future period $t^* \in T^*$. Further assume that we have extant forecasts for events $\mathbf{y}^t$ for some past period $t \in T$ that were generated from K forecasting models or teams, $M_1, M_2, \ldots, M_K$, which have prior probability $M_k \sim \pi(M_k)$. The PDF for $\mathbf{y}^t$ is $p(\mathbf{y}^t|M_k)$, the predictive PDF for the quantity of interest is $p(\mathbf{y}^{t^*}|M_k)$, the conditional probability for each model is $p(M_k|\mathbf{y}^t) = p(\mathbf{y}^t|M_k)\pi(M_k)/\sum\limits_{k=1}^{K} p(\mathbf{y}^t|M_k)\pi(M_k)$ and the and the marginal predictive PDF is $p(\mathbf{y}^{t^*}) = \sum\limits_{k=1}^{K} p(\mathbf{y}^{t^*}|M_k)p(M_k|\mathbf{y}^t)$. This can be viewed as the weighted average of the component PDFs where the weights are determined by each model's performance within the already-observed period T.

2.1 Dynamic ensemble forecasting

The EBMA procedure assumes the prior construction of multiple forecasting models or heuristics in some training period T'. The goal is to estimate the parameters for the ensemble prediction model using $\mathbf{f}_k^t$ for some period T. t is then possible to generate true ensemble forecasts ($\mathbf{f}_k^{t^*}$) for observations in the test period $t^* \in T^*$. Past applications have typically statistically post-process the predictions for out-of-sample bias reduction and treat these re-calibrated predictions as a component model. However, in the presence of sparse data we treat these raw predictions as a component model in the steps below. Additional discussion of this is below.

As a running example, let us assume that we have K forecasting efforts for modeling insurgencies in a set of countries S ongoing throughout the training (T') validation (T) and test (T^*) periods. We will associate each component forecast with a component PDF, $g_k(\mathbf{y}|\mathbf{f}_k^{s|t,t^*})$, which may be the original prediction from the forecast model or the bias-corrected forecast.

The EBMA PDF is then a finite mixture of the K component PDFs, denoted

$$p(\mathbf{y}|\mathbf{f}_1^{s|t}, \ldots, \mathbf{f}_K^{s|t}) = \sum_{k=1}^{K} w_k g_k(\mathbf{y}|\mathbf{f}_k^{s|t}), \tag{1}$$

The w_k's $\in [0, 1]$ are model probabilities and $\sum_{k=1}^{K} w_k = 1$. Roughly speaking, they are associated with each component model's predictive performance in the validation period controlling for the degree to which they offer unique insight (i.e., a model's predictions are distinct from those of other component models). We provide additional discussion about the model weights in the election forecasting example below. Details for parameter estimation are provided in Appendix A. The ensemble PDF for an insurgency in the test period t^* in country s is then:

$$p(y|f_1^{s|t^*}, \ldots, f_K^{s|t^*}) = \sum_{k=1}^{K} w_k g_k(y|f_k^{s|t^*}). \tag{2}$$

Raftery et al. (2005) propose approximating the conditional PDF as a normal distribution centered at a linear transformation of the individual forecast, $g_k(\mathbf{y}|\mathbf{f}_k^{s|t}) = N(a_{k0} + a_{k1}\mathbf{f}_k^{s|t}, \sigma^2)$. Prior applications have found that this adjustment of the component models' forecasts reduces over-fitting and improves the performance of the final ensemble forecasting model (Raftery et al. 2005).[1] Using (1) and (2) above, the EBMA PDF is then

$$p(\mathbf{y}|\mathbf{f}_1^{s|t}, \ldots, \mathbf{f}_K^{s|t}) = \sum_{k=1}^{K} w_k N(a_{k0} + a_{k1}\mathbf{f}_k^{s|t}, \sigma^2), \tag{3}$$

and the predictive distribution for some observation y is

$$p(y|f_1^{s|t^*}, \ldots, f_K^{s|t^*}) = \sum_{k=1}^{K} w_k N(a_{k0} + a_{k1}f_k^{s|t^*}, \sigma^2) \tag{4}$$

Thus, the predictive PDF is a mixture of K normal distributions each of whose mean is determined by the component prediction ($f_k^{s|t^*}$) and whose "height" (i.e, the total area under the curve for component k) is determined by the model weight w_k.

2.2 Estimation

2.3 Missing forecast components

3 Small sample adjustment

4 Applications

4.1 Predicting unemployment

4.2 Predicting presidential elections

5 Discussion

References

Raftery, Adrian E., Tilmann Gneiting, Fadoua Balabdaoui and Michael Polakowski. 2005. "Using Bayesian Model Averaging to Calibrate Forecast Ensembles." *Monthly Weather Review* 133(5):1155–1174.

Sloughter, J. McLean, Adrian E. Raftery, Tilmann Gneiting and Chris Fraley. 2007. "Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging." *Monthly Weather Review* 135(9):3209–3220.

[1] Our adjustments to the basic EBMA method for application to dichotomous outcomes, as well as details of parameter estimation, are shown in Appendix A.