Combining model forecast intervals

Given prediction percentiles, p_{ij}, from different models, M_i, for the same state of interest, y, the aim is to infer a set of percentiles, q_{cj}, that describe a combined forecast.

The percentiles of each model provide an incomplete description of the underlying predictive distribution, $p(y|M_i)$. Motivated by the work of Wallis (2005) [1], the combined forecast is attained by approximation of the mixture distribution, $\sum_i \omega_i \, p(y|M_i)$, where the weights, $\sum_i \omega_i = 1$, are assumed for now to be equal[1]. The main advantage of estimating a mixture distribution rather than directly combining the percentiles, is that the resulting estimates can be correctly interpreted as percentiles.

Initially, it is further assumed that the form of $p(y|M_i)$ is skew-normal[2]. This allows the evident non-symmetry in the individual model predictions to be captured. The validity of this assumption has been briefly and partially tested by fitting to outputs from a simple SEIR model. With each individual model predictive distribution approximated by fitting to the percentiles, the combined predictions can been estimated empirically by sampling the mixture distribution.

Detecting outliers

For each predicted output, the median, q_{c50}, of the combined distribution is estimated. The model, M_k, whose median prediction, q_{k50} lies furthest from this combined median is investigated as a potential outlier. This is decided by checking whether $q_{k_{50}}$lies within the 90% interval for the combined predictions from all models other than M_k. If it does not, the prediction is labelled an outlier. With the prediction from model M_k removed, the process is iterated, to successively identify outliers. The reported combined percentiles are formed from non-outlier predictions, while outliers are visualized as points overlaid on the bounded line charts that describe the combined prediction.

Note that if all the model predictions are sufficiently separated from one another, it is possible to label every prediction an outlier. However, in this case it can be argued that combining forecasts would be misleading. This situation could be avoided if necessary by defining a tolerance for differences in median predictions which reflects a resolution at which differences in predictions are not of interest to the decision maker.

References

1. Wallis, K. F. (2005). Combining density and interval forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics*, *67*, 983-994.
2. Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*(3), 917-1007.
3. Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359-378.

[1] Dstl are currently developing a methodology that exploits proper scoring rules [2, 3] to learn the optimal mixture weights.

[2] Dstl is also exploring non-parametric approaches to estimating the uncertain underlying distributions. This would be greatly aided by additional percentiles, or even Monte Carlo simulations from the actual models.