Global Climate Model Tracking using Geospatial Neighborhoods

Scott McQuade & Claire Monteleoni
The George Washington University, Department of Computer Science
801 22nd Street NW, Washington, DC 20052
mcquade@gwmail.gwu.edu, cmontel@gwu.edu

Climate models are complex systems of interacting mathematical models designed by meteorologists, geophysicists, and climate scientists, and run as computer simulations, to predict climate. These General Circulation Models (GCMs) simulate processes in the atmosphere and ocean such as cloud formation, rainfall, wind, ocean currents, radiative transfer through the atmosphere etc., and their simulations are used to make climate forecasts. Due to the high variance among the predictions (projections in climate science terminology) of the approximately 20 global climate models that inform the Intergovernmental Panel on Climate Change, climate scientists are currently interested in methods to combine the predictions of this multi-model ensemble, of GCMs drawn from major laboratories around the world.

Previous work provided techniques to combine the predictions of the multi-model ensemble, at various geographic scales, when considering each geospatial region as an independent problem [4, 3]. However since climate patterns can vary significantly and concurrently across the globe, this assumption is unrealistic, and could therefore limit the performance of these previous approaches. We consider a richer modeling framework in which the GCM predictions are made at higher geospatial resolutions, and our algorithm models neighborhood influences among geospatial regions.

Overview

The Tracking Climate Models (TCM) algorithm was introduced in [3] to dynamically combine the temperature predictions of the multi-model ensemble using Learn-α, an algorithm for online learning with experts under non-stationary observations [2]. The algorithm is a hierarchical learner, with updates derived as Bayesian updates of a set of generalized Hidden Markov Models (HMMs), in which the identity of the current best climate model is the hidden variable.

In this work, we propose a new algorithm, Neighborhood-augmented Tracking Climate Models (NTCM), that extends the TCM algorithm to operate in a setting in which the global climate models output predictions at higher spatial resolution. We design a variant of the algorithm to take into account regional neighborhood influences when performing updates. NTCM differs from TCM in two main ways.

1. The Learn-α algorithm is modified to include influence from a geospatial region's neighbors, in updating the weights over experts (the multi-model ensemble of GCMs' predictions in that geospatial region). This influence is parameterized by β; when $\beta = 0$ the new algorithm reduces to Learn-α, when $\beta = 1$ the maximum neighborhood influence is applied.
2. Our master algorithm runs multiple instances of this modified Learn-α algorithm simultaneously, each on a different geospatial region r, and uses their predictions to make a combined global prediction.

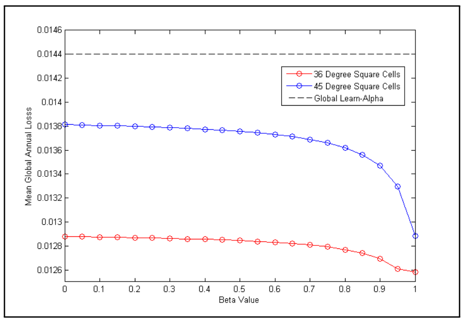
For each region r, the transition matrix among experts is defined by Equation (1). $S(r)$ is the set of all geographical regions that are spatial neighbors for the region r. This set is determined by the neighborhood scheme, which could be defined using a variety of shapes and sizes. In our experiments, we use a simple neighborhood scheme where the four immediately adjacent regions (above, below, left and right) are the possible neighbors. β is a parameter regulating the magnitude of the spatial influence, $P_{\texttt{expert}}(i, t, s)$ (conditioned over all α values) is the current probability of expert (climate model) i, as determined by the modified Learn-α algorithm for spatial neighbor s, and Z is a normalization factor that ensures that each row of the transition matrix sum to 1 (i.e. the off-diagonal terms of each row sum to α).

Experiments

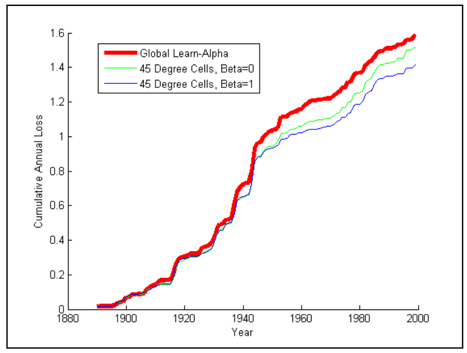
We ran experiments with our algorithm on historical data, using temperature observations, and GCM hindcasts. Historical climate model data was obtained from the International Panel on Climate Changes (IPCC) Phase 3 Coupled Model

$$P(i \mid k; \alpha, \beta, r, t) = \begin{cases} (1-\alpha) & \text{if i=k} \\ \frac{1}{Z}\left[(1-\beta) + \beta \frac{1}{|S(r)|} \sum_{s \in S(r)} P_{\texttt{expert}}(i,t,s)\right] & \text{if i}\neq\text{k} \end{cases} \tag{1}$$

$$\text{where } \; Z = \frac{1}{\alpha} \sum_{\substack{i \,\in\, \{1 \cdots n\} \\ \text{s.t. } i \neq k}} \left[(1-\beta) + \beta \frac{1}{|S(r)|} \sum_{s \in S(r)} P_{\texttt{expert}}(i,t,s)\right]$$



(a) The performance across different β values for 45 degree and 36 degree square cells, compared to the Global Learn-α performance



(b) Cumulative annual losses for the 45 degree square cells, compared to the Global Learn-α cumulative loss

Figure 1: Global results for NCTM

Intercomparison Project (CMIP3) archive [1]. Data from the Climate of the 20th Century Experiment (20C3M) was used. We ran multiple experiments using different spatial resolutions for the regions (cells) and different β values. Figure 1(a) shows the mean annual losses for a range of β values for 45 degree and 36 degree square cells. These curves show a clear trend of decreased loss with increased β values, indicating that increase influence from the spatial neighbors improves the performance. For comparison, the Global Learn-α performance is also shown on the graph.

Figure 1(b) shows how the global anomalies from the 45 degree square regions performed over time versus global Learn-α (as in the original TCM algorithm) through a graph of the cumulative annual loss. This graph indicates that for most years, and in particular for years later in the time-sequence, the NTCM cumulative global loss was less than the Learn-α loss, and with the NTCM results, the $\beta = 1$ loss was less than the $\beta = 0$ loss. The graph also shows a short period around 1940 where the cumulative losses increased sharply, with the Learn-α loss increasing significantly more than the NTCM losses.

The NTCM algorithm showed promising signs when incorporating neighborhood influence. The results showed that runs using the full neighborhood influence ($\beta = 1$) outperformed runs without neighborhood influence ($\beta = 0$) for predicting global temperature anomalies. Also at the global level, the NTCM algorithm outperformed previous global algorithms such as the Learn-α method of TCM.

References

[1] CMIP3. The World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset. `http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php`, 2007.

[2] Claire Monteleoni and Tommi Jaakkola. Online learning of non-stationary sequences. In *NIPS '03: Advances in Neural Information Processing Systems 16*, 2003.

[3] Claire Monteleoni, Gavin Schmidt, Shailesh Saroha, and Eva Asplund. Tracking climate models. *Statistical Analysis and Data Mining: Special Issue on Best of CIDU*, 4(4):72–392, 2011.

[4] C. Reifen and R. Toumi. Climate projections: Past performance no guarantee of future skill? *Geophys. Res. Lett.*, 36, 2009.

Presenter: Scott McQuade
Category: Other Applications
Secondary Category: Learning Algorithms