5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

MULTI-TASK LEARNING FROM A SINGLE TASK: CAN DIFFERENT FORECAST PERIODS BE USED TO IMPROVE EACH OTHER?

Scott McQuade[1], Claire Monteleoni[1]
mcquade@gwmail.gwu.edu, cmontel@gwu.edu

***Abstract*—We approach the problem of adaptively combining the predictions of an ensemble of seasonal climate models as a Multi-task Learning (MTL) problem. Unlike the traditional MTL setting, we only have a single functional task (combining the predictions ensemble members), where we consider multiple forecast periods from the same suite of models as our multiple learning tasks. Even though the same models generate the predictions in our "multiple tasks," we demonstrate that knowledge transfer between these forecast periods can improve ensemble predictions of the sea surface temperature in the Niño 3.4 region.**

I. BACKGROUND

The problem of combining climate model predictions can be treated as an expert tracking problem in the online setting as in [1], where an algorithm maintains a set of weights over the experts (here the climate models are the experts). The *Hedge Algorithm* [2] (also called *Static Expert* [3] when dealing with expert advice) is a common machine learning method for maintaining a set of weights over experts in the online setting via multiplicative weight updates. There has been recent work in applying Hedge and several variations to climate model ensembles [1], [4].

The goal of Multi-task Learning (MTL) is to learn multiple related tasks simultaneously, where knowledge can be transferred between tasks to improve the learning process [5]. The relatedness between tasks is frequently captured using a *Similarity Matrix* [6]. In this work we treat the problems of combining climate model predictions at different forecast periods as our multiple tasks in the MTL context.

II. APPROACH

We consider predictions from 9 different forecast periods as our multiple tasks and use the following 9-by-9 similarity matrix S:

$$\begin{bmatrix} \frac{1}{1+s} & \frac{s}{1+s} & 0 & 0 & \dots & \\ 0 & \frac{s}{1+2s} & \frac{1}{1+2s} & \frac{s}{1+2s} & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \\ & & & & & \frac{1}{1+s} \end{bmatrix}$$

Each element $S(i,j)$ represents the level of "similarity" between tasks i and j. In this simplified similarity matrix, we assume that only immediately neighboring tasks (i.e. forecast periods that differ by 1 month) are related, and the similarity is governed by the s parameter. Each row of the matrix is normalized to one so that we preserve the Hedge learning rate for a fair comparison. We modify the MTL framework proposed in [6] to handle multiplicative updates to obtain the following MTL update rule for Hedge:

$$w_{t,j}(i) = \frac{1}{Z_{t-1,j}} w_{t-1,j}(i) e^{-\sum_k S(j,k) L_{t-1,k}(i)} \quad (1)$$

Where j is the index of the task we are learning, k is an index over all tasks, S is the task similarity matrix, $L_{t-1,k}(i)$ is the loss suffered by expert i at the previous time iteration, and $Z_{t-1,j}$ is a normalization factor that ensures $w_{t,j}$ sums to 1. We use the squared loss for all loss calculations. Note that at time t, we can only evaluate losses for forecasts initiated at an appropriate amount of time into the past. For example, at time t we calculate the losses of 5.5 month forecasts initiated 5.5 months ago.

III. EXPERIMENTAL RESULTS AND DISCUSSION

We obtained monthly Sea Surface Temperature (SST) climate prediction and observation data, averaged over the Niño 3.4 region (5S to 5N, 120W to 170W), from the IRI/LDEO Climate Data Library [7]. We drew data from 6 North American Multi-Model Ensemble (NMME) groups, averaging together all available runs from each group. We ran our experiments over a hindcast from 1982 to 2010 with temperature anomalies calculated in two separate blocks: pre-October 1998 and post-October 1998, as per common practice in climate science.

[1] The George Washington University

5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

TABLE I: Summary of MTL improvements.

Forecast Period	Hedge Loss	Max MTL Improvement	s^{opt}	s^{max}
0.5 Months	0.0233	0.83%	0.625	2.125
1.5 Months	0.0871	6.35%	5.375	>50
2.5 Months	0.1465	4.01%	>50	>50
3.5 Months	0.2075	10.77%	3.875	>50
4.5 Months	0.2558	6.88%	1.750	>50
5.5 Months	0.3022	4.87%	>50	>50
6.5 Months	0.3315	5.17%	>50	>50
7.5 Months	0.3520	3.84%	>50	>50
8.5 Months	0.4111	8.39%	>50	>50

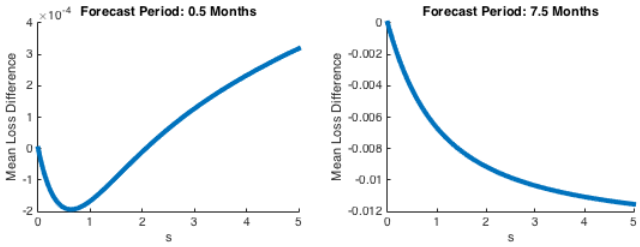

Fig. 1: The difference in mean squared error between Hedge and Hedge with MTL for 0.5 and 7.5 month forecast periods. Lower values indicate better relative performance by MTL, and values below zero indicate the MTL is outperforming Hedge.

We compared the results of Hedge and Hedge with MTL over different values for the s parameter (for task similarity) between 0 and 50. Table I summarizes the results from our experiment. s^{opt} is the s value that produced the greatest performance improvement with MTL (>50 indicates that we were continuing to see additional improvement as s increased past 50). With all forecast periods we observed improvements as we increased s from 0; s^{max} was the smallest value of s where we observed MTL performing worse than Hedge (note that this only happened with the 0.5 month period).

Figure 1 shows the difference in mean squared loss versus s values for two forecast periods.

We offer the following observations about our results:

- In all forecast periods performance improved as we increased s from 0. For all forecast periods other than 0.5 months, we did not find any s values (up to 50) that resulted in worse performance for MTL.
- While the 0.5 month period saw the smallest improvement, this is still notable since in this case all other forecasts were initiated prior to the 0.5 month forecast. Our results indicate that there was useful information from these longer forecast periods, despite their predictions being more "stale."
- For all forecast periods other than 0.5 months, we observed s^{opt} values greater than one. This indicates that we can achieve better performance by giving more weight to the losses from other forecast periods. One possible explanation for this result is that for these longer forecast periods, the losses from shorter forecast periods were more useful since they used "fresher" predictions.

IV. FUTURE DIRECTIONS

In this work we demonstrated how MTL with different forecast periods can improve the performance of a basic online learning algorithm. There are several possible extensions improve upon this technique:

1) Learning the optimal s parameter from the data.
2) Other structures for the similarity matrix could be explored, or the entire matrix could be learned from the data (as in [6], [8]).
3) Integrating our MTL technique into algorithms that are designed to handle scenarios where the "best expert" can change over time, such as [1], [3], [4].
4) Integrating other climate model outputs, such as predictions in nearby regions [4] and predictions about other climate variables (e.g. precipitation), as additional tasks.
5) Applying the idea of MTL from different forecast periods to other domains.

ACKNOWLEDGMENTS

We acknowledge Timothy DelSole, Kathleen Pegion, and Michael Tippett for their extensive discussions and advice about seasonal climate data.

REFERENCES

[1] C. Monteleoni, G. Schmidt, S. Saroha, and E. Asplund, "Tracking climate models," *Statistical Analysis and Data Mining: Special Issue on Best of CIDU*, vol. 4, no. 4, pp. 72–392, 2011.

[2] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[3] M. Herbster and M. K. Warmuth, "Tracking the best expert," *Machine Learning*, vol. 32, pp. 151–178, 1998.

[4] S. McQuade and C. Monteleoni, "Global climate model tracking using geospatial neighborhoods," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada*, pp. 335–341, 2012.

[5] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[6] A. Saha, P. Rai, S. Venkatasubramanian, and H. Daume, "Online learning of multiple tasks and their relationships," in *International Conference on Artificial Intelligence and Statistics*, pp. 643–651, 2011.

[7] "IRI/LDEO Climate Data Library." http://iridl.ldeo.columbia.edu/.

[8] A. R. Goncalves, P. Das, S. Chatterjee, V. Sivakumar, F. J. Von Zuben, and A. Banerjee, "Multi-task sparse structure learning," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 451–460, ACM, 2014.