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

AN EXAMINATION OF DEEP LEARNING FOR EXTREME CLIMATE PATTERN ANALYSIS

Gilberto Iglesias[1], David C. Kale[2], Yan Liu[2]

Abstract—**The growing volume and detail of digital climate data offer opportunities for better understanding climate and weather phenomena, but the size and complex nature of these data pose many challenges to traditional statistical learning. In this abstract, we present a preliminary examination on whether and how recent advances in *deep learning* can capture the complex interactions between climate factors and help make accurate predictions on extreme weather events, such as heatwaves.**

I. MOTIVATION

The rapid growth in the volume and detail of digital climate data means that computational methods play a key role in helping scientists to model, understand, and predict climate-related phenomena. However, climate data pose significant challenges since they are highly complex, creating nonlinear dependencies between measurements over time, and with unobserved variables.

Recent advances in *deep learning* [1] has led to state-of-the-art results in several domains with complex, nonlinear prediction functions, such as speech recognition [2], computer vision [3] and other domains. Therefore it is intriguing to explore whether deep learning approaches can effectively capture the latent structures (e.g., manifolds or clusters) in the climate data, disentangle latent factors of variation, and model sparsity in dependencies.

In this paper, we make an initial investigation and apply neural networks to predict heatwaves from longitudinal time series of climate factors. Our results suggest that neural networks can provide promising directions to improving the performance of forecasting extreme weather events. We also show some meaningful interpretations of the latent space representations discovered by neural networks.

II. METHODS

We represent a multivariate time series of P climate variables measured at T regular intervals as a matrix $\mathbf{X} \in \mathbb{R}^{P\times T}$. A *feature map* is a function $f : \mathbb{R}^{P\times T} \mapsto \mathbb{R}^D$ that maps $\mathbf{X}$ to a vector of features $\mathbf{x} \in \mathbb{R}^D$ useful for tasks like regression. The simplest feature map just flattens $\mathbf{X}$. We want to predict K different future climate and weather events, which we can represent as a vector $\mathbf{y} = [y_k]_{k=1}^K$, with features $\mathbf{x}$ as our input. To do this, we use data to learn a function $g_k(\mathbf{x}; \boldsymbol{\beta}_k)$, parameterized by $\boldsymbol{\beta}_k$, to predict y_k. When y_k is real-valued (e.g., next month's temperature), this is a regression problem; when y_k is binary (e.g., a heatweave next month), this is a classification problem. Two simple models include linear and logistic regression, respectively.

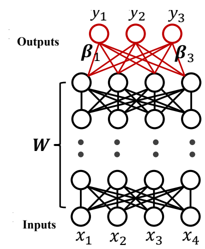

Fig. 1. Multi-task neural network for predicting climate events.

We apply a *multitask neural network* (MTNN) [4] to this problem. As illustrated in **Figure 1**, a MTNN is defined by a series of fully connected hidden layers, where the activations at level ℓ are a function of the previous layer, $\mathbf{h}_\ell = h(\mathbf{h}_{\ell-1}; \mathbf{W}_\ell)$, parameterized by a weight matrix. The lowest hidden layer $\mathbf{h}_1$ is connected to the input, while the highest is connected to K outputs. It is easy to see that each hidden layer comprises a feature map, while each output y_k is a prediction. Thus, we can use a MTNN to simultaneously learn a data-driven feature map and K different prediction models.

In our experiments, we use a MTNN with two layers of 1000 and 100 hidden units each and four outputs (two regression, two classification, explained below). We use *rectified linear activations* ($\mathbf{h}_\ell = \max(0, \mathbf{W}_\ell \mathbf{h}_{\ell-1})$) [5] and *dropout*-based (with $p = 0.25$) regularization [6]. We train the neural network using stochastic gradient descent for 2500 epochs, with minibatches of 1024, and using Nesterov momentum [7].

III. EXPERIMENTS AND DISCUSSION

We performed experiments using the climate factors data set first described in [8].[1] The data include time

Corresponding author: Y Liu, University of Southern California, Los Angeles CA, yanliu.cs@usc.edu [1] Instituto Tecnológico Autónomo de México, Mexico City, Mexico [2] University of Southern California, Los Angeles, CA

[1]Available online: http://www-bcf.usc.edu/∼liu32/data.html.

Table 1: Overall prediction performance

	Baseline	MTNN	Improve.
MSE (Tasks 1,2)	0.265 ± 0.021	0.079 ± 0.005	70.04%
AUC (Tasks 3,4)	0.806 ± 0.015	0.917 ± 0.015	13.77%

series of 18 variables related to solar radiation and atmospheric conditions (e.g., aerosol index), temperature, and weather (e.g., frost days), measured on a monthly basis from 1990 to 2002 at 125 different U.S. locations. We are interested in predicting monthly *heatwaves*, which we define as the monthly maximum temperature exceeding the five-year return level for that month. We approximate the return level using the maximum monthly temperature over the previous five years. We use a sliding window to extract all six month segments (with an overlap of three months) from each location's time series and predict: (1) the next month's maximum temperature; (2) the last (sixth) month's future maximum temperature over the next five years; (3) the presence or absence of a heatwave next month; and (4) the presence or absence of a heatwave in the last month over the next five years.

Tasks 1 and 3 involve predicting a heatwave in the immediate future. Tasks 2 and 4 involve predicting a heatwave within the next five years. The first two tasks are regressions, for which we measure performance with mean squared error (MSE). The last two are classifications, for which we measure performance with Area Under the ROC Curve (AUC). We use five folds cross validation by location to estimate performance.

Table 1 compares the average performance of MTNN against linear and logistic regression baselines. We can see that MTNN reduces overall MSE by 70% and improves overall AUC by 11%. To show that the MTNN can learn interpretable features, we first used backward variable selection to identify the second-layer feature that best predicts the next month's temperature. Then we identified the 50 input patterns that maximize that feature's activation and plot the mean and standard deviation of each variable. We can see in **Figure 2** that increasing levels of solar radiation indicates higher future temperatures.

IV. Conclusions

In this paper, we present preliminary experimental results demonstrating the potential of deep learning for analyzing complex climate data. Further experimental evaluation will be conducted for close examination so that we can provide interpretations why deep networks are appropriate models for cliamte applications.

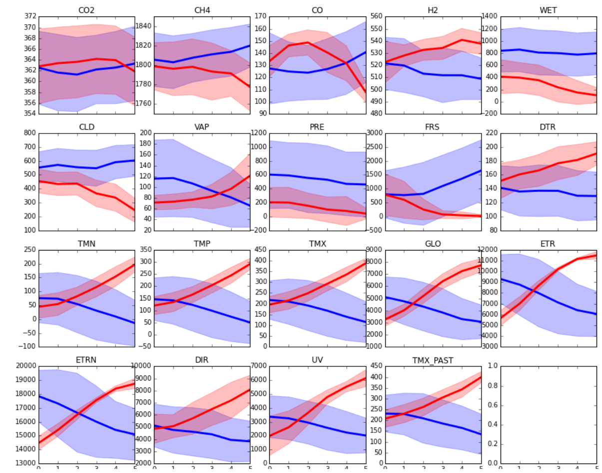

Fig. 2. Red: six-month climate pattern that predicts a higher temperature in the next month: climbing levels of solar radiation and steadily increasing temperature.

Acknowledgments

The work is partly supported by NSF IIS-1134990 and IIS-1254206 awards, Consejo at Nacional de Ciencia y Tecnología (CONACYT), Asociación Mexicana de Cultura A.C, and Alfred E. Mann Innovation Foundation. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agency, or the U.S. Government.

References

[1] Y. Bengio, A. Courville, and P. Vincent, "*Representation Learning: A Review and New Perspectives*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.

[2] D. Yu and L. Deng, "Recurrent neural networks and related models," in *Automatic Speech Recognition*, Springer, 2015.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," June 2015.

[4] R. Caruana, "Multitask learning: A knowledge-based source of inductive bias," in *Proceedings of the Tenth International Conference on Machine Learning*, ICML 10, pp. 41–48, 1993.

[5] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *International Conference on Artificial Intelligence and Statistics*, pp. 315–323, 2011.

[6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[7] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th international conference on machine learning*, ICML 30, pp. 1139–1147, 2013.

[8] A. C. Lozano, H. Li, A. Niculescu-Mizil, Y. Liu, C. Perlich, J. Hosking, and N. Abe, "Spatial-temporal causal modeling for climate change attribution," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pp. 587–596, ACM, 2009.