Convolutional Neural Networks for Climate Downscaling

Ranjini Swaminathan[+], Mohan Sridharan[*], Katharine Hayhoe[+]
Climate Science Center[+], Department of Computer Science[*], Texas Tech University

1 Introduction

A key challenge in climate modeling is the assessment of the impact of global climate variables on regional weather measurements such as temperature and precipitation. This assessment is usually done by *downscaling* the output of (coarse resolution) global climate models to regional (high resolution) predictions. There are two independent downscaling pathways: *dynamic* and *statistical*. Dynamic downscaling uses high resolution physical models of regional climate in conjunction with global models to make regional climate predictions. Although dynamic downscaling methods account for regional geographic variations, they are computationally expensive even on modern supercomputers. Statistical downscaling algorithms, on the other hand, establish correlations between global model outputs and historical (decadal) weather observations, developing statistical models that translate global model output into regional weather projections.

In recent times, many machine learning (ML) algorithms and statistical methods have been used for statistical downscaling, e.g., Bayesian frameworks, artificial neural networks, support vector machines, multilayer neural networks [4, 5, 7, 9] and ensemble methods (e.g., boosting, bagging and stacking).However, climate downscaling poses some open challenges such as: incomplete representation (and non-linearity) of physical processes in global models, non-stationarity in the global models and (hence) the relationships with regional observations, and computational complexity of processing large Petabyte-scale datasets. Many existing ML algorithms are limited in their ability to model complex functions with many variations that represent *deep* relationships between input and output variables [1].

Research in ML shows that architectures designed to model complex functions need to be of sufficient *depth* (i.e., maximum length of path from any input to any output) to prevent poor generalization [2]. Many ML algorithms do *not* correspond to deep architectures, e.g., under certain assumptions, decision trees have two levels, logistic regression has one, kernel machines have two levels and ensemble methods add one level to the base learner [1]. Even in multilayered neural networks, learning algorithms perform poorly when complex functions require (error) gradients to be propagated across many levels. Another limitation of many ML algorithms is the use of *local estimators*, i.e., input space is partitioned into regions and the degrees of freedom in the underlying model are used to account for variations of the target function in these regions. As a result, algorithms such as decision-trees and local kernel machines (e.g., Gaussian processes and manifold learning algorithms) may not generalize well for complex functions with many variations [3].

2 Technical Approach

We propose to use *deep architectures* [1] to learn the complex relationships between global climate model output and observed station data. Deep architectures use multiple levels of *distributed representations*, where the input pattern is represented by features that are not mutually exclusive [1]. We start by investigating the use of convolutional neural networks (CNNs), a benchmark for visual recognition tasks [6]. A CNN is a multilayered network of *neurons*, a deep architecture motivated by the structure of the human visual system. A CNN has layers of two types: *convolutional* and *subsampling*. These layers have *topographic structure*, i.e., neurons or nodes have a 2D spatial position (e.g., in an image) in addition to a receptive field that influences the response of the corresponding neurons. Neurons in each layer (other than the first layer) are associated through weights with neurons in specific positions in a previous layer. Such a structure allows for better propagation of the error gradients over space (and time) in a network with many layers. In addition, the hierarchical connections help capture local variability and enable modeling of prior knowledge of the connectivity. Since learning parameters for all layers is rather challenging, unsupervised learning is used within each layer to move parameters in the right direction [8]. Unsupervised learning also decomposes the overall learning problem into subproblems with different levels of abstraction, e.g. features extracted in one layer serve as input features to the next higher layer.

We hypothesized that CNNs will result in robust models of the complex relationships between global climate model output and observed historical data because:

- The input parameters are not treated as independent variables but as arising from some spatial structure whose topology is captured by the model.
- CNNs provide the flexibility of using a larger number of layers (e.g., in comparison to multilayered neural nets), automatically varying the number of layers to provide better generalization.

The hypotheses are evaluated experimentally in the context of decadal climate downscaling datasets. The results are compared with those obtained with ensemble methods and multilayered neural nets. These results will serve to quantify the uncertainty in the data sources and to (potentially) guide the use of deep architectures for climate modeling.

References

[1] Y. Bengio. Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

[2] Y. Bengio. Learning Deep Architectures for AI. Technical report, Universite de Montreal, Department of Computer Science and Operations Research, Report 1312, 2009.

[3] Y. Bengio and Y. LeCun. Scaling Learning Algorithms Towards AI. In L. Bottou and O. Chapelle and D. DeCoste and J. Weston, editor, *Large Scale Kernel Machines*. MIT Press, 2007.

[4] S. Ghosh. SVM-PGSL Coupled Approach for Statistical Downscaling to Predict Rainfall from GCM Output. *Journal of Geophysical Research*, 115:D22102, 2010.

[5] M. Khan, P. Coulibaly, and Y. Dibike. Uncertainty Analysis of Statistical Downscaling Methods. *Journal of Hydrology*, 319(1):357–382, 2006.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86:2278–2324, 1998.

[7] J. Murphy, B. Booth, M. Collins, G. Harris, D. Sexton, and M. Webb. A Methodology for Probabilistic Predictions of Regional Climate Change from Perturbed Physics Ensembles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365:1993–2028, 2007.

[8] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun. Efficient Learning of Sparse Representations with an Energy-based Model. In *Advances in Neural Information Processing Algorithms*, 2006.

[9] B. Tomassetti, M. Verdecchia, and F. Giorgi. NN5: A Neural Network-based Approach for the Downscaling of Precipitation Fields-Model Description and Preliminary Results. *Journal of Hydrology*, 367(1-2):14–26, 2009.