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

DISCOVERING SPATIAL AND TEMPORAL PATTERNS IN CLIMATE DATA USING DEEP LEARNING

Charles Anderson[1], Imme Ebert-Uphoff[2], Yi Deng[3], Melinda Ryan[1]

***Abstract*—Recent experiments involving the training of artificial neural networks with multiple layers, sometimes referred to as deep learning, have demonstrated the ability to automatically identify features that are critical to solving complex pattern classification tasks, such as speech recognition. Similar to speech, atmospheric data sets often consist of multiple time series with unknown, complex interrelationships. In this project we seek to explore what kind of interrelationships can be discovered in climate data by applying the framework of artificial neural networks. As a first application we look at establishing relationships between top of atmosphere radiative flux and air/surface temperatures. This is an important application, since a thorough understanding of those relationships is essential for understanding the effect of CO_2-induced warming on the Earth's energy balance and future climate. We describe the basic idea, first observations and plans for future work.**

I. INTRODUCTION

Artificial neural networks (ANNs), first developed in the mid 1900's as models of the brain, are robust function approximators consisting of layers of simple computational units with weighted interconnections. Recently many-layered, or deep, ANNs have been used to solve difficult problems in image and speech recognition, for example [1]. Fig. 1 shows the typical structure of such a network, consisting of input units (on the left), that are connected to one or more layers of hidden units, which are connected to a layer of output units.

In this project we explore what we can learn by applying ANNs to climate data. The type of ANN used here is called an "autoencoder" network—it is trained to approximate each input sample as its output. The autoencoder ANN used here, shown in Fig. 1, includes a "bottleneck layer", i.e., a layer with a small number of units that must be trained to capture as much of each sample's variation as possible in a small number of dimensions, in this case three. While this is a relatively shallow network by today's deep-learning standards, the autoencoder paradigm is often used in training deep networks. We are experimenting with deeper networks which should reveal more complex relationships among atmospheric variables.

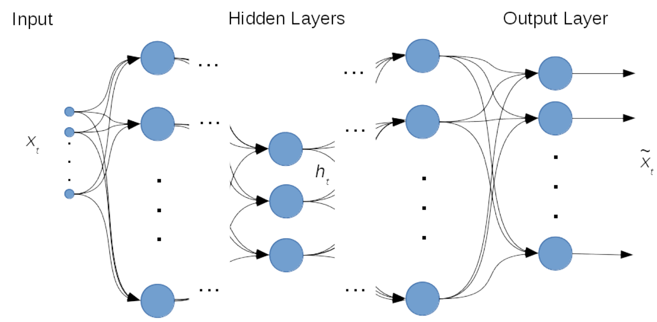

Fig. 1. Neural network including one bottleneck layer with only three units.

As a first application we try to gain insights into the complex dynamics governing the interactions between the radiative flux at the top of the atmosphere (TOA) and air/surface temperatures. Note that in another submitted abstract (submission #6) two of the authors approach the same application using the method of causal discovery. The two methods, artificial neural networks and causal discovery, are entirely different in nature and are expected to yield very different types of insights, thus complementing each other.

We are using NASA CERES data [http://ceres.larc.nasa.gov/index.php] for shortwave flux at TOA (sw), long wave flux at TOA (lw) and solar insolation at TOA (si). We are using NASA MERRA data [http://gmao.gsfc.nasa.gov/merra/] for daily air temperature at 850, 500 and 50hPa and surface temperature. We are using daily data from March 1, 2000 to Dec 31, 2013 (5,054 days), and low spatial resolution of 20 x 20 degrees.

Corresponding author: C. Anderson, anderson@cs.colostate.edu. [1]Dept. of Computer Science, Colorado State University, Fort Collins, CO, USA. [2]Dept. of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, USA. [3]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA.

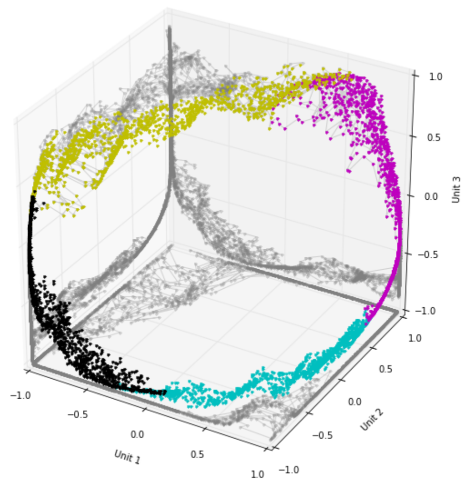

Fig. 2. The three-dimensional representations of all data points produced by the three-units in the bottleneck layer. Colors represent quarters of the year—first quarter is cyan, second is magenta, third is yellow, fourth is black. Two-dimensional projections are in gray.

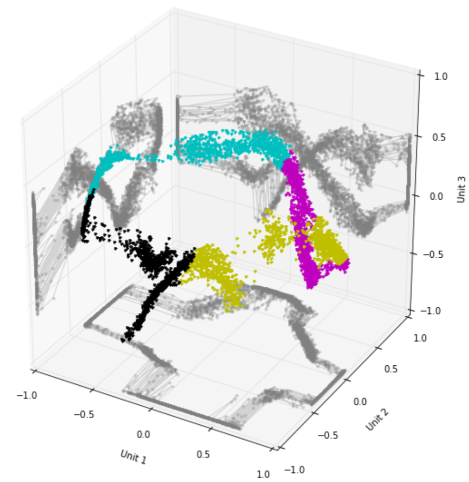

Fig. 3. Results from 2nd model. The three-dimensional representations of all data points produced by the three units in three separate bottleneck networks trained sequentially.

II. Training the ANN Model

Each sample has 756 components, consisting of six atmospheric fields at all 126 locations (7 latitudes and 18 longitudes). The output of the ANN is trained to approximate all 756 components in the mean-square sense. In this work, error is minimized using the scaled conjugate gradient algorithm [2]. Fig. 2 is a visualization of the learned three-dimensional encoding of each 756-dimensional sample. The most obvious pattern in all three dimensions is an annual cycle, made clear by coloring points from the four quarters of each year in different colors.

A model with less-correlated low-dimensional components might reveal more interesting patterns, thus we implemented a second model. To bias the bottleneck units to learn less-correlated functions, a novel approach is taken in which three different autoencoder networks are trained sequentially. First, an autoencoder network with a single unit in the bottleneck layer is trained to fit the data. Then the approximations of the samples, provided by the output of the trained autoencoder, is subtracted from the original samples and a second autoencoder network with one bottleneck unit is trained to approximate this difference. In this way, the second autoencoder network is trained to approximate the residual in the data after removing the approximation of the data given by the first autoencoder. This is repeated a third time. This sequential projection and subtraction operation is motivated by the Gram-Schmidt algorithm [3] used in singular value decomposition.

Fig. 3 shows the three-dimensional encoding that results from this sequential method of training three autoencoders. The three units represent very different projections of the data. As expected, the first unit appears to have captured the annual cycle. By removing this information from the data, the second unit is able to learn different relationships, as well as the third unit. Currently the patterns in this encoding are being analyzed in terms of which co-variations in climate variables are being represented.

III. Discussion and Future Work

The example and initial results provided here illustrate the basic approach we plan to take, namely to train neural networks from data and to then *study the network properties* to discover interesting patterns in the data. We have only just started to implement these ideas, so the initial results presented here serve as an *illustration of the general process*, and we have yet to see what kind of scientific discoveries can be made using this approach.

Acknowledgments

This work is supported by the NSF Climate and Large-Scale Dynamics (CLD) program through Grant AGS-1147601 awarded to Yi Deng, and a collaborative grant (AGS-1445956 and AGS-1445978) awarded to Yi Deng and Imme Ebert-Uphoff. Melinda Ryan is supported through an REU supplement associated with grant AGS-1445978.

References

[1] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: an overview," in *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8599–8603, May 2013.

[2] M. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks*, vol. 6, no. 4, pp. 525–533, 1993.

[3] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins University Press, fourth ed., 2012.