A Hybrid Deep Learning Approach For Chaotic Time Series Prediction Based On Unsupervised Feature Learning

Norbert Ayine Agana

Advisor: Abdollah Homaifar

Autonomous Control & Information Technology Institute (ACIT),
Departmenent of Electrical and Computer Engineering,
North Carolina A&T State University

June 16, 2017

Outline

Time Series Prediction

1. Time series prediction is a fundamental problem found in several domains including climate, finance, health, industrial applications etc
2. Time series forecasting is the process whereby past observations of the same variable are collected and analyzed to develop a model capable of describing the underlying relationship
3. The model is then used to extrapolate the time series into the future
4. Most decisions made in society are based on information obtained from time series analysis provided it is converted into knowledge

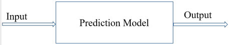

Figure 1

Time Series Prediction Models

1. Statistical methods: Autoregressive(AR) models are commonly used for time series forecasting
   1. Autoregressive(AR)
   2. Autoregressive moving average (ARMA)
   3. Autoregressive integrated moving average (ARIMA)
2. Though ARIMA is quiet flexible, its major limitation is the assumption of linearity form of the model: No nonlinear patterns can be captured by ARIMA
3. Real-world time series such as weather variables (drought, rainfall, etc.), financial series etc. exhibit non-linear behavior
4. Neural networks have shown great promise over the last two decades in modeling nonlinear time series
   1. Generalization ability and flexibility: No assumptions of model has to be made
   2. Ability to capture both deterministic and random features makes it ideal for modeling chaotic systems
5. Nonconvex optimization issues occurs when two or more hidden layers are required for highly complex phenomena

Problem Statement

1. Deep neural networks trained using back-propagation perform worst than shallow networks
2. A solution is to initially use a local unsupervised criterion to (pre)train each layer in turn
3. The aim of the unsupervised pre-training is to:
    - obtain useful higher-level representation from the lower-level representation output
    - obtain better weights initialization

Motivation

1. Availability of large data from various domains(Weather, stock markets,health records,industries etc.)
2. Advancements in hardware as well in machine learning algorithms
3. Great success in domains such as speech recognition, image classification, computer vision
4. Deep learning applications in time series prediction, especially climate data, is relatively new and has rarely been explored
5. Climate data is highly complex and hard to model, therefore a non-linear model is beneficial
6. A large set of features have influence on climate variables

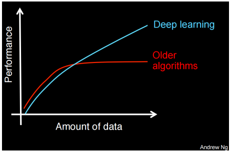

Figure 2: How Data Science Techniques Scale with Amount of Data

Deep Learning

1. Deep learning is an artificial neural network with several hidden layers
2. There are a set of algorithms that are used for training deep neural networks
3. Deep learning algorithms seek to discover good features that best represent the problem, rather than just a way to combine them

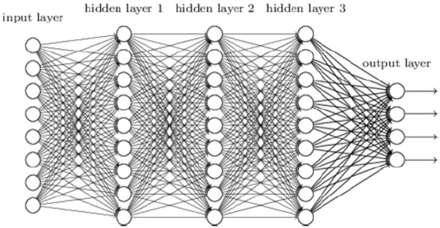

Figure 3: A Deep Neural Network

Unsupervised Feature Learning and Deep Learning

1. Unsupervised feature learning are widely used to learn better representations of the input data
2. The two common methods are the autoencoders(AE) and restricted Boltzmann machines(RBM)

Stacked Autoencoders

1. The stacked autoencoder (SAE) model is a stack of autoencoders
2. It uses autoencoders as building blocks to create a deep network
3. An autoencoder is a NN that attempts to reproduce its input: The target output is the input of the model

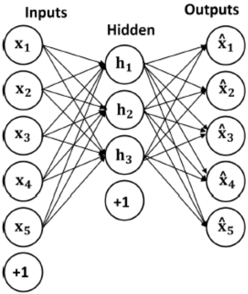

Figure 4: An Example of an Autoencoder

Deep Belief Networks

1. A Deep Belief Network (DBN) is a multilayer neural network constructed by stacking several Restricted Boltzmann Machines(RBM)[3]
2. An RBM is an unsupervised learning model that is learned using contrastive divergence

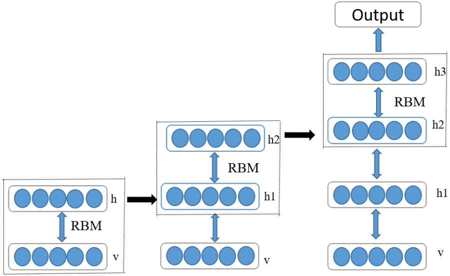

Figure 5: Construction of a DBN

Proposed Deep Learning Approach

1. We propose an empirical mode decomposition based Deep Belief Network with two Restricted Boltzmann Machines
2. The purpose of the decomposition is to simplify the forecasting process

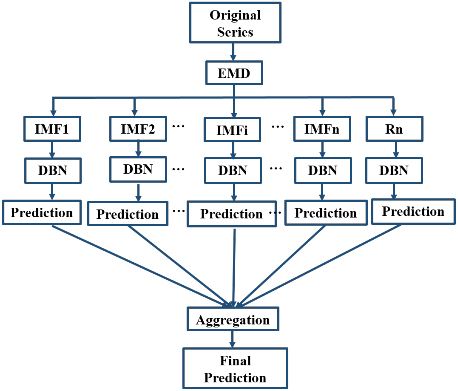

Figure 6: Flowchart of the proposed model

Proposed Deep Learning Approach

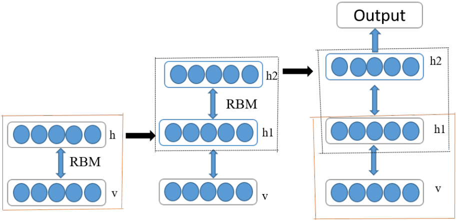

Figure 7: Proposed Model

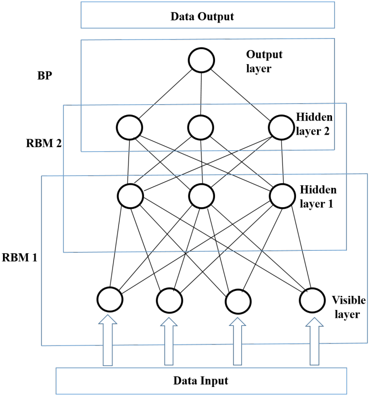

Figure 8: DBN with two RBMs

Restricted Boltzmann Machines (RBMs) I

1. An RBM is a stochastic generative model that consists of only two bipartite layers: visible layer v and hidden layer h
2. It uses only input(training set) for learning
3. A type of unsupervised learning neural network that can extract meaningful features of the input data set which are more useful for learning
4. It is normally defined in terms of the energy of configuration between the visible units and hidden units

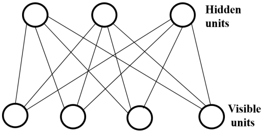

Figure 9: An RBM

Restricted Boltzmann Machines (RBMs) II

The joint probability of the configuration is given by [4]:

$$P(v, h) = \frac{e^{-E(v,h)}}{Z},$$

Where Z is the partition function (normalization factor):
$Z = \sum_{v,h} e^{-E(v,h)}$
and $E(v, h)$, the energy of configuration:

$$E(v, h) = -\sum_{i=visible} a_i v_i - \sum_{j=hidden} b_j h_j - \sum_{ij} v_i h_j w_{ij}$$

Training of RBMs consists of sampling the h_j given v (or the v_i given h) using Contrastive Divergence.

Training an RBM

1. Set initial states to the training data set (visible units)
2. Sample in a back and forth process

Positive phase: $P(h_j = 1|v) = \sigma\left(c_j + \sum w_{ij} v_i\right)$
Negative phase: $P(v_i = 1|h) = \sigma\left(b_i + \sum w_{ij} h_j\right)$

3. Update all the hidden units in parallel starting with visible units, reconstruct visible units from the hidden units, and finally update the hidden units again

$$\triangle w_{ij} = \alpha\left(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}\right)$$

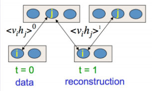

Figure 10: Single step of Contrastive Divergence

4. Repeat with all training examples

Deep Belief Network

A Deep belief network is constructed by stacking multiple RBMs together. Training a DBN is simply the layer-wise training of the stacked RBMs:

1. Train the first layer using the input data only (unsupervised)
2. Freeze the first layer parameters and train the second layer using the output of the first layer as the input
3. Use the outputs of the second layer as inputs to the last layer (supervised) and train the last supervised layer
4. Unfreeze all weights and fine tune the entire network using error back propagation in a supervised manner.

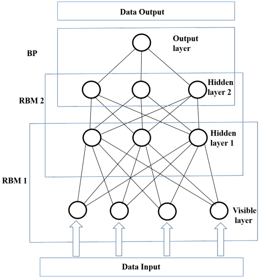

Figure 11: A DBN with two RBMs

Empirical Mode Decomposition (EMD)

1. EMD is an adaptive data pre-processing method suitable for non-stationary and nonlinear time series data [5]
2. Based on the assumption that any dataset consists of different simple intrinsic modes of oscillations
3. Given a data set, $x(t)$, the EMD method will decompose the dataset into several independent intrinsic mode functions (IMFs) with a corresponding residue, which represents trend using the equation[6]:
   $X(t) = \sum_{j=1}^{n} c_j + r_n$
   where the c_j are the IMF components and r_n is a residual component

The Hybrid EMD-BBN Model

1. A hybrid model consisting of Empirical Mode Decomposition and a Deep Belief Network (EMD-DBN) is proposed in this work

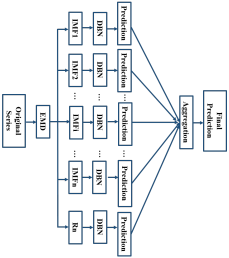

Figure 12: Flowchart of the hybrid EMD-DBN model

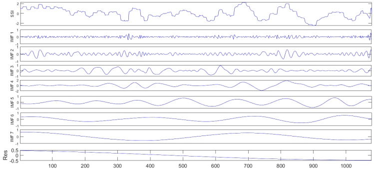

Figure 13: EMD decomposition of SSI series: The top is the original signal, followed by 7 IMFs and the residue

Summary of the proposed approach

The following few steps are used [1],[2]:

1. Given a time series data, determine if it is nonstationary or nonlinear
2. If yes, decompose the data into a fine number of IMFs and a residue using the EMD
3. Divide the data into training and testing data (usually 80% for training and 20% for testing)
4. For each IMF and residue, construct one training matrix as the input for one DBN. The input to the DBN are the past five observations
5. Select the appropriate model structure and initialize the parameters of the DBN. Two hidden layers are used in this work
6. Using the training data, pre-train the DBN through unsupervised learning for each IMF and the residue
7. Fine-tune the parameters of the entire network using the back-propagation algorithm
8. perform predictions with the trained model using the test data
9. Combine all the prediction results by summation to obtain the final output

Prediction of Solar Activity

1. The solar activity is characterized, among others, by means of the relative sunspot number
2. Sunspots are dark spots that are often seen on the suns surface.
3. Its known to influence several geophysical processes on earth
4. For example, atmospheric motion, climate anomaly, ocean change, etc. all have different degree of relation with sunspot number
5. It is also a good determiner for solar power generation
6. Due to the complexity of the sunspot number change, modeling methods have encountered troubles trying to describe its change rules

Description of Data

1. The monthly time series representing the solar activity cycle during the last 268 years is used.
2. The data represents the sunspot number, that is an estimation of the number of individual sunspots from 1949 to 2016: A total of 3216 observations

- We used monthly sunspot time series for the years 1749-1960 as the training set, and 1961-2016 for cross-validation (testing)

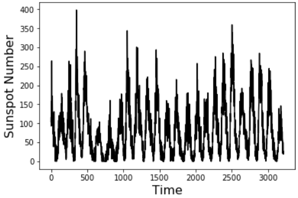

Figure 14: Monthly Total Sunspot Number: 1749 - 2016

Decomposition of Sunspot Number Series

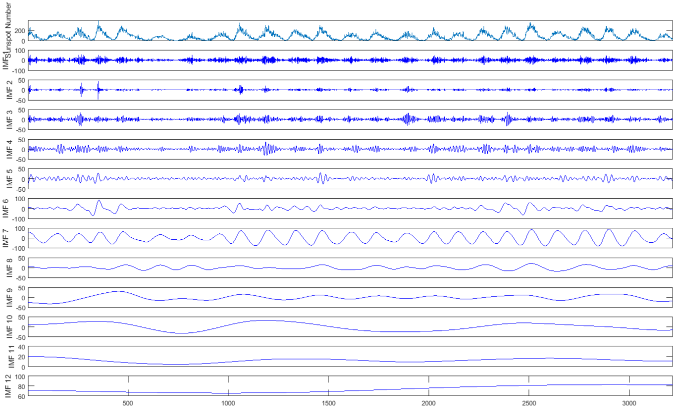

Figure 15: EMD Decomposition

Results and Discussion

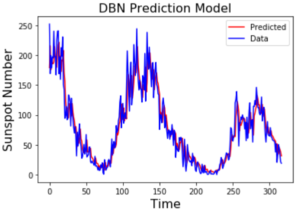

Figure 16: DBN prediction Results

Table 1: Prediction Errors

Model	MSE	RMSE	MAE
MLP(5 10 1)	0.00359	0.05992	0.04798
DBN(5 10 10 1)	0.00345	0.05865	0.04396
EMD-MLP(5 10 1)	0.00078	0.09205	0.02101
EMD-DBN(5 10 10 1)	0.00020	0.01438	0.01070

Application to Drought Prediction

1. Drought is a natural disaster that occurs with great impact on society
2. Occurs when there is a significant deficit in rainfall compared to the long-term average
3. Affects water resources, agricultural and socioeconomic activities
4. Drought prediction is very vital in limiting their effects
5. Predictions can be useful in the control and management of water resources systems and mitigation of economic, environmental and social impacts

Figure 17

Study Area and Data

1. The case study is carried out using data from the Gunnison River Basin, located in the Upper Colorado River Basin with a total drainage area of $5400 km^2$
2. Monthly Streamflow observations from 1912 to 2013 are used
3. Standardized Streamflow Indices (SSI) are calculated based on the streamflow data

Figure 18: Location of the Gunnison River Basin [7]

Summary of the Proposed Model

1. Obtain the different time scale SSI
2. Decompose the time series data into several IMFs and a residue using EMD
3. Divide the data into training and testing data
4. Pre-train each layer bottom up by considering each pair of layers as an RBM
5. Finetune the entire network using the back-propagation algorithm
6. Use the test data to test the trained model

Table 2: Prediction Errors for SSI 12

Model	MSE	RMSE	MAE
MLP(5 10 1)	0.00422	0.06468	0.04211
EMD-MLP(5 10 1)	0.00209	0.03580	0.02882
DBN(5 10 10 1)	0.00211	0.04593	0.02852
EMD-DBN(5 10 10 1)	0.00131	0.02257	0.01649

Table 3: Prediction Errors for SSI 24

Model	MSE	RMSE	MAE
MLP(5 10 1)	0.00303	0.05507	0.03969
EMD-MLP(5 10 1)	0.00179	0.04399	0.04249
DBN(5 10 10 1)	0.00125	0.03535	0.01876
EMD-DBN(5 10 10 1)	0.00077	0.02780	0.01009

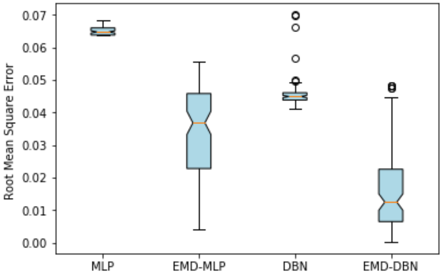

Figure 19: Comparison of the RMSE for SSI 12 forecast

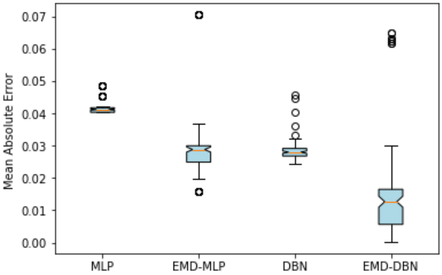

Figure 20: Comparison of the MAE for SSI 12 forecast

Conclusion

1. This study explored a deep belief network for drought prediction. We proposed a hybrid model comprising of empirical mode decomposition and deep belief network (EMD-DBN) for long term drought prediction
2. The results of the proposed approach are compared with both DBN, MLP and also EMD-MLP.
3. Overall, the hybrid EMD-DBN model was found to provide better forecasting results for SSI 12 and SSI 24 in the Gunnison River Basin
4. Performance of both MLP and DBN improved when the drought time series are decomposed

Summary of Contributions and Future Work I

Contributions:

1. Constructed a DBN model for time series prediction by adding a final layer which simply map learned features with the target
2. Improved the performance of the proposed model by integrating empirical mode decomposition to form a hybrid EMD-DBN model
3. Calculated Standardized drought indices using the generalized extreme value (GEV) distribution instead of a gamma distribution
4. Applied the proposed model to drought prediction

Futurework:

1. Optimize the structure of the model(e.g. number of hidden layers, hidden layer size, and also learning rate) by using search methods such as grid search or random search
2. Use a linear neural network to aggregate the individual predictions instead of just summing them

Summary of Contributions and Future Work II

3. Predict extreme precipitation indices across the Southeastern US
4. Apply the model to predict other climate variables such as precipitation and temperature using satellite images

Thank You

?

References I

[1] Norbert A Agana and Abdollah Homaifar.
A deep learning based approach for long-term drought prediction.
In *SoutheastCon, 2017*, pages 1–8. IEEE, 2017.

[2] Norbert A Agana and Abdollah Homaifar.
A hybrid deep belief network for long-term drought prediction.
In *Workshop on Mining Big Data in Climate and Environment (MBDCE 2017), 17th SIAM International Conference on Data Mining (SDM 2017)*, pages 1–8, April 27 - 29, 2017, Houston, Texas.

[3] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh.
A fast learning algorithm for deep belief nets.
Neural computation, 18(7):1527–1554, 2006.

[4] Geoffrey Hinton.
A practical guide to training restricted boltzmann machines.
Momentum, 9(1):926, 2010.

References II

[5] Zhaohua Wu, Norden E Huang, Steven R Long, and Chung-Kang Peng.
On the trend, detrending, and variability of nonlinear and nonstationary time series.
Proceedings of the National Academy of Sciences, 104(38):14889–14894, 2007.

[6] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu.
The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis.
In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995. The Royal Society, 1998.

References III

[7] Shahrbanou Madadgar and Hamid Moradkhani.
A bayesian framework for probabilistic seasonal drought forecasting.
Journal of Hydrometeorology, 14(6):1685–1705, 2013.