Energy Load Forecasting using Deep Learning Approach LSTM and GRU in Spark Cluster

Introduction

Since the energy wastage is a threat to sustainability, making the process of energy generation and consumption for future, efficient is extremely crucial. Based on the interval of time period, energy load forecasting can be grouped into three categories: (1) Short-term-load - forecasting (STLF) which is usually from 1 hour to 1 week (2) Medium-term-load forecasting (MTLF), ranging from 1 week to 1 year, and (3) Long-term load forecasting (LTLF) which is longer than one year. Various models have been developed for realizing forecasting accuracy such as Regression, Statistical and state space methods. AI based approaches have been explained based on Expert Systems, Fuzzy Logic Systems, Artificial Neural Networks(ANNs). Hybrid approaches to Time-Series analysis utilizing ANN are not uncommon, presents a model for Time-Series forecasting using ANN and ARIMA models. Electric load forecasting is primarily a discrete & univariate time series, many statistical time series models can be applied for electric load forecasting such as Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA) models and lot of their variants. For more information on time series and deep learning approach to it. The main focus of the presented paper is to solve Electric load forecasting problem using deep learning approach on a Spark Cluster. Deep learning allows models composed of multiple layers to learn pattern representations within the data.

Gated Recurrent Unit

The GRU is a variant of the LSTM and was introduced by K, Cho. The GRU was inspired by the LSTM unit but is considered simpler to compute and implement. It retains the LSTM's resistance to the vanishing gradient problem, but its internal structure is simpler and therefore easier to train since fewer computations are needed to make update to its hidden state. The GRU has an update gate & a reset gate similar to the forget & input gates in the LSTM unit. The update gate defines how much previous memory to keep around and the reset gate defines how to combine the new input with the previous memory. The major difference is that the GRU fully exposes its memory content using only integration (but with an adaptive time constant controlled by the update gate). These are the mathematical functions used in controlling the gating mechanism in GRU cell:-

$$z = \sigma(W_z h_{t-1} + U_z x_t)$$
$$r = \sigma(W_r h_{t-1} + U_r x_t)$$

$$c = tanh(W_c(h_{t-1} \otimes r) + U_c x_t)$$
$$h_t = (z \otimes c) + \left((1 - z) \otimes h_{t-1}\right)$$

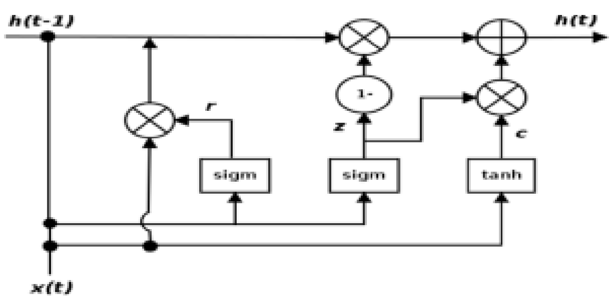

Figure 4. *A Gated recurrent unit cell*

Experimental setup

In the presented paper we use Spark and a cluster of 7 machines of core i5 processor with 4 GB primary memory to improve deep learning based model's performance .The models selected for experimental analysis include LSTM and GRU. Distributing the computations among the 7-node cluster, we were able to train 7 different configuration of models concurrently.

Dataset & Experimental Results

This Section deals about the dataset chosen for testing and training the models discussed earlier in the presented paper using Spark and a Cluster of machines .It then discusses the results obtained for the two models with different configurations investigated both on Sequential machine and on Spark Cluster.

Dataset

The selected methods were implemented on a dataset of electric power consumption in household with a one -minute sampling rate .The data set contained power consumption measurements collected between mid-December 2006 to mid-April 2008 .The First 14 months data was chosen to train both the model and the last two month data was used as testing. To improve the performance of deep neural models hyperparameter tuning was done firstly on single machine chosen 1 configuration at a time. Then each different configuration was trained parallel on an individual node. Using Spark for computation we find the best set of hyper-parameters for RNN, LSTM & GRU training, resulting in 6X reduced training time and 23% lower error rate.

Experimental results

GRU Network

The above mentioned configurations of deep neural nets were implemented with the standard GRU network to predict few steps ahead. The results shows that GRU model performed much better than the standard LSTM network architecture, and hence give less error ratios with both training & test dataset. In RNN, LSTM and GRU, the configuration with 3 Hidden Layers and 30 units in each hidden unit's promises to give the least RMSE 0.592, 0.584 & 0.562 respectively. The best configuration with least error value was chosen for testing purpose. Deploying models: Apply the selected trained RNN/LSTM and GRU model for forecasting on a large amount of new data on a Spark Cluster

Conclusions

The goal of the presented work was to investigate the effectiveness in using LSTM and GRU based neural network architectures for energy load forecasting. Both of them were trained and tested on a Spark Cluster. The accuracy of forecasting load with the initial set of hyper-parameters is 99.1%. The best chosen configuration of LSTM & GRU has 99.39% accuracy on the test set, which is around 34% reduced test error. We were able to train 7 models concurrently using a 7 - node cluster, which is 6x faster compared to training time on a single node sequential machine. GRU model was better at forecasting load as compared to LSTM.