An Efficient Electricity Generation Forecasting System Using Artificial Neural Network Approach with Big Data

Kiran Kumar ,Nikita Prakash ,Sreenag Hoskote Pranesh,Zuqiang Ke

The United States has a large geographical diversity with enormous amount of power consumption in the world.Due to lack of centralized control, there is a large imbalance of power consumption/generation.This results in wastage of large quantity of electricity in few states where generation exceeds the consumption,whereas the other states are suffering from insufficient power. The statistical data also predicts the same.In order to reduce the gap, power generation in the United States needs a centralized control to predict the power generation bases based on the demand/consumption.

There is a large amount of noise and high volatility in datasets therefore prediction/forecasting is always a problem.Machine learning algorithm fails to handle such huge datasets.Therefore, we need a new technology to search,analyze and format the data before we use the Artificial Neural Network to deal with the statistical dataset. Big Data technology is one of the best choices to deal with the above problem.

The framework consists of two parts: first, raw data were processed and made into a suitable format; then the data is fed into an Artificial Neural Network (ANN) model for training. To deploy our work, we collect power generation data of all the states of the U.S. in the past, and store it in a distributed database. By this procedure, we hope we can come up with a linear relation between power generation and demand, which can be easily apply to the centralized control. We choose a Big Data product such as Hadoop and MapReduce to deal with the large dataset and Artificial Neural Network (ANN) to generate the results for predicting power demand.

Data is stored in multiple computer nodes with Hadoop Distributed File System (HDFS) technology, and MapReduce be used to allocate assignment and to handle large datasets. Then manipulated data is extracted from each computing node in the desired format.This data is used on ANN for training and predict the future power generation and consumption. In this cluster there is one master node that works with several slave nodes and does the nodes management and data computation. Once the computation is done, the results are submitted to the master and stored in HDFS. The normalization factor is calculated based on Equation. Normalized value of each column data (Vnor) is:

$$V_{nor} = Value \times NF$$

Where, Normalization factor, NF=Vmax/F,Vmax= Maximum value of the column, F= Converted the value of Vmax to floating point value.

The data obtained will be sent to Backpropagation Neural Network (BPNN) for training.The obtained data is divided into two sets: 90% used for training the network and rest 10% for testing the network. There 12 nodes in input layer, 6 nodes in hidden layer and 1 in the output layer.In each node we have activation function, usually triggered after certain level of input. The activation function for each node is as following:

$$f = \frac{1}{1+e^{-(net\ input)}}$$

The BP algorithm has a few problems, such as local minima and overtraining, but it is still a good choice due to its performance, and once it is successfully trained, it has the ability to detect pattern with huge noise. Using Big Data technology with ANN is one of the best choices to solve the problem that has a huge amount of noisy dataset.The end results show us it is a

feasible and effective way to yield a close proximity between the forecasted and actual power generation values.