**American Journal of Information Science and Computer Engineering**
Vol. 5, No. 2, 2019, pp. 77-81
http://www.aiscience.org/journal/ajisce
ISSN: 2381-7488 (Print); ISSN: 2381-7496 (Online)

# An Overview of Big Data and Machine Learning for Energy Forecasting

**Farook Sattar***

Department of Electrical & Computer, Engineering, University of Victoria, Victoria, Canada

## Abstract

Energy forecasting is a technique to predict future energy needs to achieve demand and supply equilibrium. This paper presents an overview on Big Data and machine learning technology in the context of energy forecasting. The overall objective of Big Data is to discover useful information and knowledge that might otherwise be overlooked or discounted. On the other hand, machine learning helps make complex energy systems more efficient as the systems can learn from a large volume of collected data, detect regular patterns, and optimize its own operations. The energy forecasting plays a vital role to predict energy consumption for large commercial customers. The electrical energy sector is now looking for ways to put a higher level of accuracy and reliability into forecasting electrical loads for the next day and for the next week. It is further complicated by the fact that this sector is dependent on other energy sectors including wind, solar, gas, and hydro, all of which are directly affected by weather events. Therefore, development of intelligent prediction systems based on Big Data as well as machine learning techniques is an emerging issue to pursue a sophisticated and highly tuned decision support system for energy forecasting and it involves integration of various large volume of data from numerical energy prediction models, statistical datasets, real time observations, and human intelligence to optimize forecasts for low-cost energy generation.

## Keywords

Big Data, Energy Forecasting, Machine Learning

Received: April 21, 2019 / Accepted: June 16, 2019 / Published online: June 24, 2019



## 1. Introduction

Big Data (BD) phenomenon has emerged as a result of vast amounts of data that are becoming available across a wide range of application domains across science, business and government. Research on big data and machine learning will be necessary for serving scientists, engineers, educators, citizens and decision makers who have unprecedented amounts and types of data available to them.

BD is composed of text, image, video, audio, mobile or other forms of data collected from multiple datasets, and is rapidly growing in size and complexity. It has created a huge volume of multidimensional data within a very short time period. This raises several new challenges, including; how to classify BD for multiple datasets, how to analyze BD for different forms of data, and how to visualize BD without the loss of information. Large datasets allow for the development of better prediction models able to simulate energy systems at different levels. In the context of demand response, BD plays a pivotal role, as it allows electric utilities to allow a large amount of information and develop models as well as simulation tools that can guide the development of such initiatives [1].

BD shows high potential for energy forecasting [2]. Partly, the credit goes to the roll-out of smart metering and the development of smart appliances, driving the Internet of Things and generating large volume of data compared to the analog meter device. This is reinforced by data availability of

* Corresponding author
E-mail address: farook_sattar@yahoo.com.sg

other data types such as machine-to-machine transaction logs, GPS data, social media data and increasingly accurate weather data including their distributions per point in time. As a result, this enables utility companies to obtain far more knowledge about their customers consumption patterns.

In energy industry, a true demonstration of the potential of BD is in developing more realistic forecasts of energy generation, for example by wind farms and fleets. A more accurate forecast of energy generation allows a utility to optimize operations. Specially, they can more closely match generation with demand and thus requires smaller reserves to be on standby. A good example of this is the work being done by National Center for Atmospheric Research and Xcel Energy [3, 4]. The two organizations have teamed up to use artificial-based software and sophisticated weather modeling techniques to analyze turbine wind speed and energy generation measurements and produce wind energy generation forecast. The forecasts have improved since more data has been incorporated into the analysis. For example, earlier the forecasts used data from just one or two weather stations per wind farm. Now the center collects information from nearly every wind turbine. As a result, the accuracy has improved significantly and the forecasts saved Xcel nearly as much as money as in the three previous years combined.

The machine learning (ML), on the other hand, opens up significant opportunities for improving and refining forecast and prediction models, allowing managers to make decisions supported by data. Predictive analysis can help utilities to manage electricity production, transmission, distribution, supply and demand. They are capable of managing nondispatchable loads (such as renewable), sending the proper guiding signals to several independent customers, accommodating changing weather and grid conditions in real-time, etc.

# 2. The Role of ML for Energy Forecasting

Various machine learning techniques have been used to forecast electrical energy needs, including neural networks (NN) [5], support vector machines (SVM) [6]. Among them neural networks have been widely used for electricity forecasting and considered them suitable for industrial energy forecasting [7]. On the other hand, support vector regression (SVR) is considered as an emerging technique for energy forecasting.

Neural networks (NN) are a family of machine learning models inspired by the human brain and used to approximate functions that are generally known. Like a human brain, neural networks consist of interconnected neurons. There are many types of neural networks such as radial basis function networks, Kohonen self-organizing networks, and recurrent networks; however, here the focus is on feed-forward neural networks (FFNN) because the FFNN is one of the most frequently used NNs used for energy forecasting [8].

In the following, we briefly introduce the two machine learning techniques – FFNN and SVR – extensively used in energy prediction:

## 2.1. Feed-Forward Neural Networks (FFNN)

Figure 1 shows a three-layer FFNN that can be used to approximate non-linear functions without assuming relationships between inputs and outputs. The information in the FFNN moves in one direction, from the input layer through the hidden layers) to the output. In such a network, there are no connections between the neurons in the same layer. The number of neurons in the input layer is equal to the number of input features, and the number of neurons in the output layer is equal to the number of outputs. A FFNN can have more than one hidden layers and the number of neurons in each hidden layer are chosen by the user.

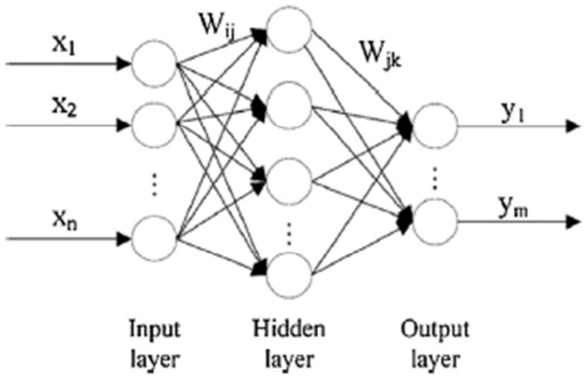


**Figure 1.** Feed Forward Neural Networks [8].

The output of each neuron in the hidden layer is determined as follows:

$$y_j = \varphi\left(\sum_{i=1}^{n} w_{ij}x_i + w_{io}\right) \quad (1)$$

where the $\mathbf{x}_i$ are neuron inputs, the $\mathbf{w}_{ij}$ are synaptic weights connecting the i-th neuron in the input layer to the j-th neuron in the hidden layer, and $\mathbf{w}_{ij}$ is a bias which shifts the decision boundary, but does not depend on any inputs. $\varphi$ is an activation function which is usually modelled as a step or sigmoid function. The output of the neurons in the output layer is modelled in the same way, with the weights corresponding to connections between the hidden and output layers.

FFNN weights are learned during the training phase, using back propagation in conjunction with an optimization method

such as gradient descent. To start the learning process, the weights are randomly initialized. Next, the input is applied and the output is calculated according to the feed forward process described earlier. The calculated output is then compared to the known output, and the calculated error is propagated backwards through the network. During this back propagation, the weights are adjusted according to the optimization method to reduce the error for that specific input. The process is repeated for all training examples, and the overall process is repeated until the error drops below a pre-defined threshold.

### 2.2. Support Vector Regression (SVR)

Support vector machines (SVM) are supervised learning models used for classification and regression problems; a version of SVM for regression is referred to as support regression (SVR). SVR is characterized by a high degree of generalization, which indicates the model's ability to perform accurately on new, previously unseen data. In SVR, support vectors are training samples which lie on the $\varepsilon$-tube bounding decision surface, as illustrated in Figure 2. Observations within the $\varepsilon$-tube do not influence predictions; in other words, residuals less that $\varepsilon$ do not get penalized.

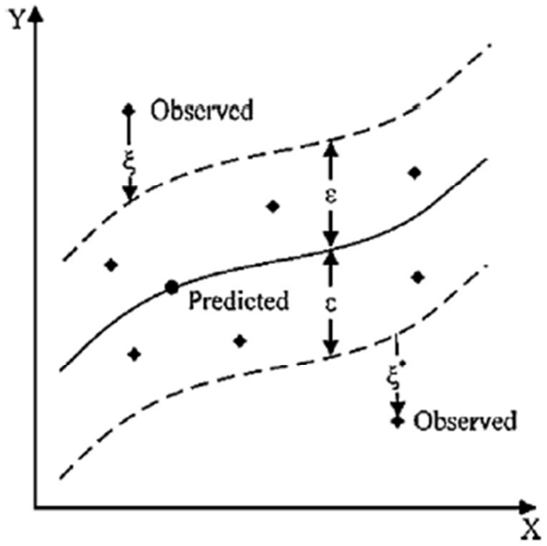


**Figure 2.** Nonlinear SVR [9].

Suppose that an output $Y$ is modelled as a function of input variables $X$, given a training data set $\{(X_i, Y_i)\}_{i=1, N}$. The SVR approximates the relationship between input and output as:

$$Y = W.\phi(X) + b \tag{2}$$

where $\phi(X)$ is a nonlinear kernel function which non-linearly maps from the input space $X$ to the feature space. Coefficients $W$ and $b$ are determined by minimizing the following function:

$$Minimize\ \frac{1}{2}\|w^2\| + C\frac{1}{N}\sum_{i=1}^{N}\xi_i + \xi_i^* \tag{3}$$

subjects to constraints:

$$Y_i - W.\phi(X_i) - b \leq \varepsilon + \xi_i \tag{4}$$

$$W.\phi(X_i) + b - Y_i \leq \varepsilon + \xi_i^* \tag{5}$$

$$\xi_i, \xi_i^* \geq 0 \tag{6}$$

where $W$ is a weight vector which needs to be as flat as possible to achieve good generalization. Terms $\xi_i$ and $\xi_i^*$ capture residuals beyond the $\varepsilon$ boundary (see Figure 2), and cost $C$ is the regularization parameter that determines the penalty for errors greater than $\varepsilon$.

The radial basis function (RBF) is a widely used kernel for mapping the input space to a high-dimensional feature space. The RBF is also efficient to compute and has only one parameter that needs to be determined; hence, this work also uses the radial basis kernel,

$$K(x, x') = \exp(-\gamma\|x - x'\|^2) \tag{7}$$

In eqn. (7), the parameter $\gamma$ specifies the influence of each data point.

## 3. The Current Focus

Here we discuss on the interesting developments in the area of deep data analytics known as deep learning, which is an emerging technique to access and manipulate really large datasets [10-12] to improve the energy prediction. Deep learning refers to a recently developed set of *generative* machine learning techniques that autonomously generate high-level representations from raw data sources, and using these representations can perform typical machine learning tasks such as classification, regression and clustering. Many of the most important deep learning techniques are extensions of neural network methods and a simple way to understand them is to think of multiple layers of neural networks linked together. Taking raw data input at the first layer the output of the next layer is a set of high level features which are passed to a further layer which in tune generates a set of higher level features. This continues for a number of layers until eventually output (e.g. a prediction) is produced.

In contrast to more conventional machine learning algorithms, deep learning has an advantage of potentially providing a solution to address the data analysis and learning problems found in massive volumes of input data. More specifically, it assists in automatically extracting complex data representing from large volumes of unsupervised data. This makes it a valuable tool for Big Data Analytics, which involves data analysis from very large collections of raw data that is generally unsupervised and uncategorized. The hierarchical learning and extraction of different levels of

complex data abstractions in deep learning provides a certain degree of simplification for Big Data analytics tasks, especially for analyzing massive volumes of data, semantic indexing, data tagging, information retrieval, and discriminative tasks such as classification and prediction [13].

The main advantage of deep learning is referred to the drastically increased chip processing abilities, the lowered cost of computing hardware, and the recent advances in machine learning. Deep Neural Networks (DNNs) are multilayer networks with many hidden layers, whose weights are fully connected and often initialized or pretrained using stacked Restricted Boltzmann Machine (RBM) or Deep Belief Networks (DBNs) [14].

The RBM is a generally probabilistic model between input units (visible), **x**, and latent units (hidden), **h** (see Figure 3).

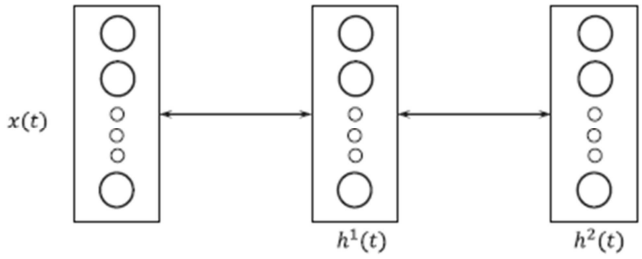


**Figure 3.** A 2-layer RBM for static data. The visible units are fully connected to the first hidden layer $h^1$[15].

The visible and hidden units are connected with a weight matrix, **W** and have bias vectors **c** and **b**, respectively. There are no connections among the visible and hidden units. The RBM can be used to model static data. The energy function and the joint distribution for a given visible and hidden vector is defined as:

$$E(\mathbf{x},\mathbf{h}) = \mathbf{h}^T\mathbf{W}\mathbf{x} + \mathbf{b}^T\mathbf{h} + \mathbf{c}^T\mathbf{v} \quad (8)$$

$$P(\mathbf{x},\mathbf{h}) = \frac{1}{Z}\exp(E(\mathbf{x},\mathbf{h})) \quad (9)$$

where $Z$ is the partition function that ensures that the distribution is normalized. For binary visible and hidden units, the probability that hidden unit $h_j$ is activated given visible vector **x** and the probability that visible unit $x_i$ is activated given hidden vector **h** are expressed as:

$$P(h_j|\mathbf{x}) = \sigma\left(b_j + \sum_i W_{ij}\, x_i\right) \quad (10)$$

$$P(x_i|\mathbf{h}) = \sigma\left(c_i + \sum_j W_{ij}\, h_j\right) \quad (11)$$

where σ (·) is the activation function. The logistic function $\sigma(x) = \frac{1}{1+e^{-x}}$ is a common choice for activation function. The parameters **W**, **b**, and **v** are trained to minimize the reconstruction error using constructive divergence. The learning rule for the RBM is:

$$\frac{\partial log\mathbf{P}(\mathbf{x})}{\partial W_{ij}} \approx < x_i h_j >_{data} - < x_i h_j >_{recon} \quad (12)$$

where $< \cdot >$ denotes the average value over all training samples.

DBN is a pretraining, unsupervised step that utilizes large amount of unlabeled training data for extracting structures and regularities in input features [16, 17]. DBN not only uses a huge amount of unlabeled training data but also provides good initialization weights for DNN. Moreover, overfitting and underfitting problems can be tackled by using the pretraining step of DBN. Unlike ANNs, the DBNs have only one visible layer composed of input and output nodes, and all the remaining layers are hidden. The DBNs model the relation between the observation vector x and the variables $h^l$ of the $l$-th hidden layer, with $l$=1,…, M by the following joint distribution:

$$P(\mathbf{x},\mathbf{h}^1,\mathbf{h}^2,\ldots,\mathbf{h}^M) = \left(\prod_{j=0}^{M-2} P(\mathbf{h}^j|\mathbf{h}^{j+1})\right) P(\mathbf{h}^{M-1}|\mathbf{h}^M) \quad (13)$$

In eqn. (13), $\mathbf{x}=\mathbf{h}^\circ$, $P(\mathbf{h}^{M-1}|\mathbf{h}^M)$ is the top layer joint distribution in the top-level RBM and the computation of probability for the conditional distribution is $P(\mathbf{h}^j|\mathbf{h}^{j+1})$ (for the j-th layer). The general structure of a DBN is presented in Figure 4.

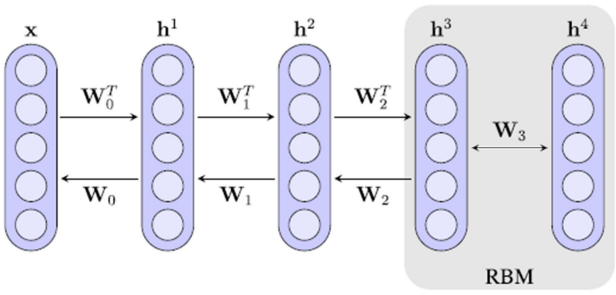


**Figure 4.** The general structure of Deep Belief Network (DBN) [18].

Due to the high complexity of the net, the training phase cannot be approached by adopting the same strategies applied for the ANNs, e.g. gradient descent or Lavenberg-Marquardt algorithm. So, a fast training method has been developed: the ``greedy algorithm'' [19]. The algorithm plans to train each hidden layer as an RBM that models the output data of the previous layer as its visible layer.

For the first hidden layer, $\mathbf{h}^1$, the input vector $\mathbf{x}=\mathbf{h}^\circ$ is assumed as a visible layer, and used to tune the equivalent weight matrix $\mathbf{W}_0$. The representation of the input data for the training of the second hidden layer, $\mathbf{h}^2$, is obtained by mapping the input vector through the weight matrix $\mathbf{W}_0^T$ ($T$ stands for transpose). The procedure is repeated for each layer, each time propagating the mapped input vector.

DNN has shown great performance in recognition and classification, and traffic flow detection [13]. However, DNN

has high computational cost and difficult to scale [20]. DSN addresses the scalability problem of DNN, simple classifiers are stacked on top of each other in order to construct more complex classifiers [21].

## 4. Conclusion

This paper provides an overview on Big Data and machine learning in the aspect of energy forecasting. Here, our aim is to introduce Big data and machine learning techniques useful for energy prediction rather than their applications. The Big Data and machine learning technology allow the information to be analyzed in more detail than with traditional technology, and the application of it to the energy sector is an innovative idea. The two most popular machine learning techniques –NN and SVR- which are extensively used for energy prediction, are presented first. Then the newly focused deep learning technique is introduced due to its vast potentiality in the predictive analysis for energy forecasting.

## Declaration of Interests

The authors declare that they have no competing interests.

## References

[1] B.-A. S.-Leech, B. Barry, M. Muratori, and B. J. Yurkovich, "Big Data issues and opportunities for electric utilities", Renewable and Sustainable Energy Reviews, 52, 937-947, 2015.

[2] K. Grolinger, A. L'Heureux, M. A. M. Capretz, and L. Seewald, "Energy forecasting for event venues: Big data and prediction accuracy', Energy and Buildings, 112, 222-233, 2016.

[3] K. Parks, Y.-H. Wan, G. Wiener, and Y. Liu, "Wind energy forecasting: A collaboration of the National Center for Atmospheric Research (NCAR) and Xcel Energy", National Renewable Energy Laboratory (NREL). Colorado, USA, 2011.

[4] C.-D. Yue, C.-C. Liu, C.-C. Tu, and T.-H. Lin, "Prediction of power generation by offshore wind farms using multiple data sources", Energies, 12 (4), 2019.

[5] F. Ardakani and M. Ardehali, "Long-term electrical energy consumption forecasting for developing and developed economics based on different optimized models and historical data types", Energy, 65, 452-461, 2014.

[6] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Yaylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy", Applied Energy, 123, 168-178, 2014.

[7] L. Suganthi and A. Samuel, "Energy models for demand forecasting – A review", Renewable and Sustainable Energy Reviews, 16 (2), 1223-1240, 2012.

[8] R. Z. Jovanovic, A. A. Sretenovic, and B. D. Zivkovic, "Ensemble of various neural networks for prediction of heating energy consumption", Energy and Buildings, 94, 189-199, 2015.

[9] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using Support Vector Regression: Investigating the temporal and spatial monitoring granularity on performance accuracy", Applied Energy, 123, 168-178, 2014.

[10] Z. H. Kilimci, A. O. Akyuz, M. Uysal, S. Akyokus, M. O. Uysal, B. A. Bulbul, and M. A. Ekmis, "An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain", Complexity (Hindawi Pub.), Volume 2019, 1-15, 2019.

[11] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, ``Deep learning applications and challenges in big data analytics'', Journal of Big Data (Springer), 2 (1), 1-21, 2015.

[12] X.-W. Chen and X. Lin, "Big data deep learning: challenges and perspectives", IEEE Access, 2, 514-525, 2014.

[13] Y. Lv., Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: a deep learning approach", IEEE Trans. Intell. Transp. Syst., 99, 1-9, 2014.

[14] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning", APSIPA Trans. Signal Inform. Process., 3, 2014.

[15] M. Langkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling", Pattern Recognition Letters, 42, 11-24, 2014.

[16] J. Schmidhuber, "Deep learning in neural networks: An overview", Neural Networks, 61, 85-117, 2015.

[17] J. Ngiam, A. Coares, A. Lahiri, B. Prochnow, Q. V. Le, and A. Y. Ng, "On optimization methods for deep learning", Int. Conf. on Machine Learning (ICML), 265-271, 2011.

[18] M. Fagiani, S. Squartini, L. Gabrielli, S. Spinsante, and F. Piazza, "A review of datasets and load forecasting techniques for smart natural gas and water grids: Analysis and experiments", Neurocomputing, 170, 448-465, 2015.

[19] Y. Bengio, P. Lamblin, D. Popovivi, and H. Larochelle, "Greedy layer-wise training of deep networks", Neural Comput., 18 (7), 1527-1554, 2006.

[20] L. Deng and D. Yu, "Deep complex net: a scalable architecture for speech pattern classification", Interspeech, 2011.

[21] K. P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.