APPLYING DEEP LEARNING WITH EXTENDED KALMAN FILTER AND GENETIC ALGORITHM OPTIMIZATION FOR WATER DISTRIBUTION DATA-DRIVEN MODELING

Zheng Yi Wu[3], Atiqur Rahman[2] and Qiao Li[1]
[1]Bentley Systems, Incorporated, Watertown, CT, USA
[2]College of William & Mary, Williamsburg, VA, USA
[3]Denver University, Denver, Colorado, USA
[1]*zheng.wu@bentley.com*

ABSTRACT

Data-driven analysis has recently emerged as an important task for smart water management as large amount of various data collected via smart meters, sensors and data loggers. Among the methods developed for data-driven modeling, deep neural network (DNN) is proved as the competitive and generic approach to solve many challenging problems, including but not limited to voice recognition, natural language processing, image classification etc. Deep belief network (DBN) is one of the DNNs and widely used for data analysis. This paper extended authors' previous research in applying DBN model with the genetic algorithm to integrate with the Extended Kalman Filter (EKF). It results in a comprehensive and generic approach, by which the genetic algorithm is employed to optimize the configuration of the DBN and the EKF is applied to assimilate the newly available data with the trained DBN model so that the model can be updated whenever new data becomes available. The proposed method has been tested in the case studies of different domains, including but not limited to water distribution systems. The results show that the deep learning method integrated with EKF has resulted in good performance.

Keywords: Deep learning, data-driven analysis, predictive modeling, optimization, smart water systems

1   INTRODUCTION

Infrastructure has been the foundation of our society. It provides transportation, power generation safe drinking water and much more. Infrastructure must be sustainable and resilient to the incidents and the deteriorations. To address the issue of aging infrastructure, sensors are placed onto infrastructure to monitor infrastructure health conditions. Ubiquitous sensing is continuously capturing data and generating large volume of data. With the ubiquitous sensing and interconnectivity, there is an imperative need for developing efficient and effective data analytics to extract useful information from big data, discover knowledge from information and acquire wisdom from knowledge to facilitate good decision-making. Intelligent data analytics play an important role at each and every step of the process.

In water distribution system, a conventional SCADA system has been widely adopted to monitor the critical facilities, such as pump stations, storage tanks, and other key control points. While SCADA systems are very well developed, but the data collected from these systems are usually accumulated but no deep analysis is conducted to make good use of the data. Over last decade, Automatic Meter Reading (AMR) and Advanced Metering Infrastructure (AMI) technologies are gaining more and more acceptance in water industry. This is primarily driven by improving

operation efficiency. Widely adopted SCADA and emerging AMR/AMI are the backbone of implementing effective monitoring programs, which rapidly produce the datasets in large size (so-called big data). Its value cannot be realized unless useful/actionable information is retrieved or extracted, and used for systematic simulation and optimization modelling to facilitate the decision making, and finally improve the automation control for triggering/communicating with the instruments in the field. However, conducting big data analysis calls for development of generic data-driven modelling tool. Based on our previous work [1], this paper proposes an enhanced deep belief network (DBN) [2] learning model with the Extended Kalman Filter (EKF) [3] that is integrated with DBN to enable assimilated learning from the given dataset. The application of EKF with the optimal parameters of the neural network results in the output covariance, which can be used to establish the confidence levels of the model prediction.

2 INTEGRATED DEEP LEARNING FRAMEWORK

An integrated deep belief learning framework has been developed with the extended Kalman filter and genetic algorithm optimization tool [4], which provides high performance computing features involved many CPUs (and/or compute nodes) and many GPUs to accelerate the model training process.

2.1 Deep belief network

Deep Belief Network (DBN) was originally proposed by Hinton, Osindero and Teh in 2006 [2]. It has been shown that the DBN can be constructed by stacking a number of Restricted Boltzmann Machines (RBMs) and training them in a greedy layer wise manner to learn more complex models than conventional ANN architectures. An RBM is a fully connected bipartite graph with two layers namely visible and hidden layers. The neurons in visible and hidden layers are designed to take binary or Gaussian values as input, so that several different architectures can be obtained with different value type, namely binary-binary, Gaussian-binary or Gaussian-Gaussian.

Based on the RBM model, the DBN model's conditional probability distribution between observed vector and hidden layers is as follows:

$$P(x, h^1, h^2, \ldots, h^l) = \left(\prod_{k=0}^{l-2} P(h^k | h^{k+1}) \right) P(h^{l-1}, h^l) \tag{1}$$

where x corresponds to the input layer and is equivalent to h^0. $P(h^{l-1}, h^l)$ is the joint probability distribution of visible units of a RBM at layer l conditioned on the hidden units of the RBM at layer $l - 1$. Figure 1 shows an example of DBN.
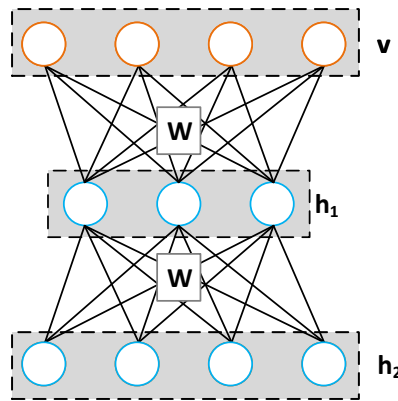
Fig.1 DBM model

For the predictive analysis application, the Gaussian-binary RBM must be implemented with the following energy function [5]:

$$E(v,h) = -\sum_{i \in V} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j \in H} b_j h_j - \sum_{i,j} \frac{v_i}{\sigma_i} h_j w_{ij} \qquad (2)$$

Where V is the set of neurons in visible layers while H designates the set of neurons in hidden layers. v_i and h_j are the states of visible units and hidden units respectively; while a_i is bias weight for the visible units and b_j is for the hidden units. w_{ij} is the weights between the neuron i and the neuron j. σ_i is the standard deviation of the Gaussian noise for visible unit i. Learning the noise associated with each visible unit poses a challenge, but it can be overcome by normalizing each component of the data though subtracting the mean and divided by the variance. The learning rate is adjusted by reducing it by one or two magnitude from what is used for binary-binary RBM training. Reduction in learning rate is required to keep the weights emanating from certain components becoming very large. Gaussian visible units with rectified linear hidden units has been employed in the Gaussian-binary RBM. In this model, the hidden units are approximated by $\max(0, N(0,1))$ where $N(0,1)$ is a Gaussian noise with zero mean and unit variance.

2.2   DBN high performance optimization

Different configurations (structures) of DBN models may result in the huge difference on the prediction performance. To achieve the good accuracy for the prediction, the best configuration of DBN must be adopted for the given problem. However, there is no general rules for choosing the best DBN model structure, a trial-and-error approach is usually employed in practice. In order to construct a robust DBN model efficiently and effectively

, the Darwin Optimization Framework [15] (DOF) is integrated with DBN to optimize the DBN configuration [12]. DOF is a genetic algorithm-based optimization software tool. It enables heterogeneous high performance optimization with many CPUs (or compute nodes), where DBN model structures and training parameters are optimized, and many GPUs, where DBN training is accelerated by using hundreds of computing threads. During the optimization, the DBN

configurations, such as the number of hidden layers and the number of neurons at each hidden layer, are encoded into binary strings, so that they can be evolved by the genetic algorithm. For each generation in the evolution, the binary strings will be decoded into the configuration information to generate the DBN models to calculate their fitness, which is the prediction accuracy. The final optimized configuration given by DOF is also decoded into the configuration information so that the DBN model can be constructed for the given problem.

2.3   Assimilated training

Unlike conventional artificial neural network, a DBN model constriction involves two-step training namely the pre-training and fine-tuning. With traditional gradient-based training methods, the corrections of the weights in the neural network are calculated with the gradients of the activation functions by chain rule. However, since the gradient of the commonly used activation function is in the range of (-1, 1), and the corrections of the later layers are the multiplications of these small gradients by chain rule, the update values for the weights of the "front layers" (the layers close to the output side) become smaller as the number of layers of the network increases, and thus the "front layers" cannot be trained effectively. This is so-called the vanishing gradient problem, consequently the DBN cannot be trained effectively with the traditional gradient-based training methods. To address the vanishing gradient challenge, the pre-training [13] is adopted to obtain the good representation of the data. It enables the neural network model to learn layer by layer, so that the "front layers" can be well trained. After the pre-training, the neural network is fine-tuned by the traditional training method, e.g. backpropagation, to further improve its accuracy. In this paper, the EKF is employed to fine-tune the neural network. The structure of the EKF integrated with DNN is shown in Figure 2 to enable the assimilated learning. The principle is to use the DBN to "mimic" the output of the physical system in real world. For prediction applications, the inputs of the real system and the DBN are the historical data. If the outputs of the real system and the DBN are the same, it means that the prediction of DBN is accurate. Otherwise, the difference between the outputs will be feed-backed to the EKF to adjust the estimation of the states, i.e. the weights and biases of the neural network. This way, the new output of the real system can be timely assimilated with the DBN via EKF.
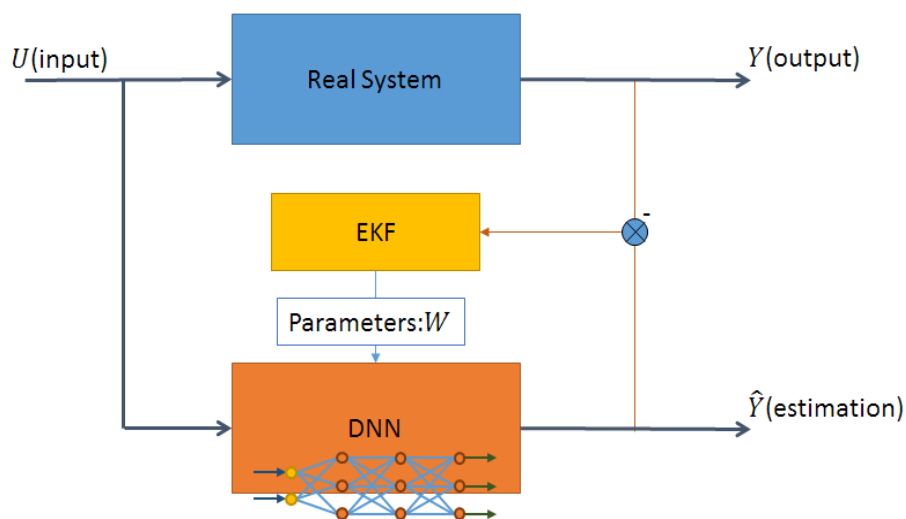

Figure 2 Assimilated training with extended Kalman filter

3   EXAMPLE APPLICATION

The integrated approach for enhanced DBN model has been applied to a few real use cases for data-driven analysis. The learning process follows Figure 2, which includes the Darwin optimization, pre-training and fine tuning with EKF. Then, the prediction is performed with the trained neural network model. The results are compared with the neural network trained with back propagation (BP) method.

3.1 Daily water consumption prediction

The first case is the water consumption prediction. The dataset is the daily water consumption data of a real water system from 2005 to 2008 [6]. In this case, a deep belief neural network is optimized with 7 inputs, 3 hidden layers (7 neurons in each layer) and 1 output is used. The inputs of the neural network are the previous 7 days' water consumption data, and the output is the prediction of the water demand for the next day. The neural network is trained with the data in 2005 and 2006, then tested for prediction of the daily water consumption of 2007 and 2008.
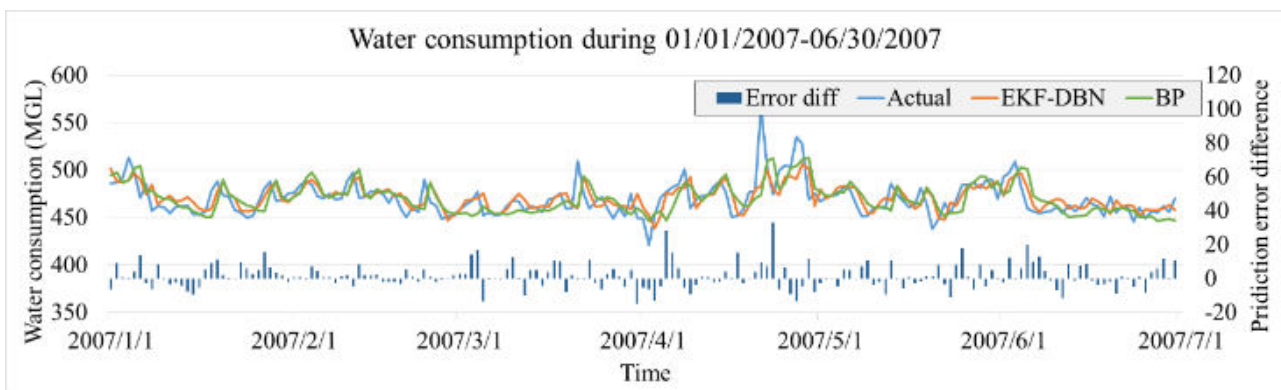

Figure 3 Comparison of prediction and the actual daily water consumption

The sample prediction results of the proposed EKF integrated DNN (EKF-DNN) is presented in Fig.6 compared with the prediction from BP trained neural network (BP-DNN) and the actual data. The line charts in the figures are the actual water consumption values and the predictions of BP and proposed DBN-EKF methods. The bar charts in the figures are the difference between the prediction errors of two methods, in which the positive value indicates the proposed DBN-EKF method has better prediction accuracy. Based on the results, the Mean Absolute Percentage Error (MAPE) is 1.998% for EKF training and 3.178% for BP training. Meanwhile, the RMSE is 6.22 for EKF-DNN and 9.21 for BP-DNN. Therefore, the EKF-DNN is more accurate than BP-DNN. Also, as shown in Figure 3, the prediction of the conventional BP method gets less accurate in 2008 than in 2007. This is because the BP trained neural network has fixed parameters that cannot keep tracking the system changes, while the EKF-DNN can be updated with the new input data. Therefore, the EKF-DNN is able to produce consistently accurate prediction for a long period of time, since it keeps learning to assimilate the new data feature along the time horizon while being used for prediction.
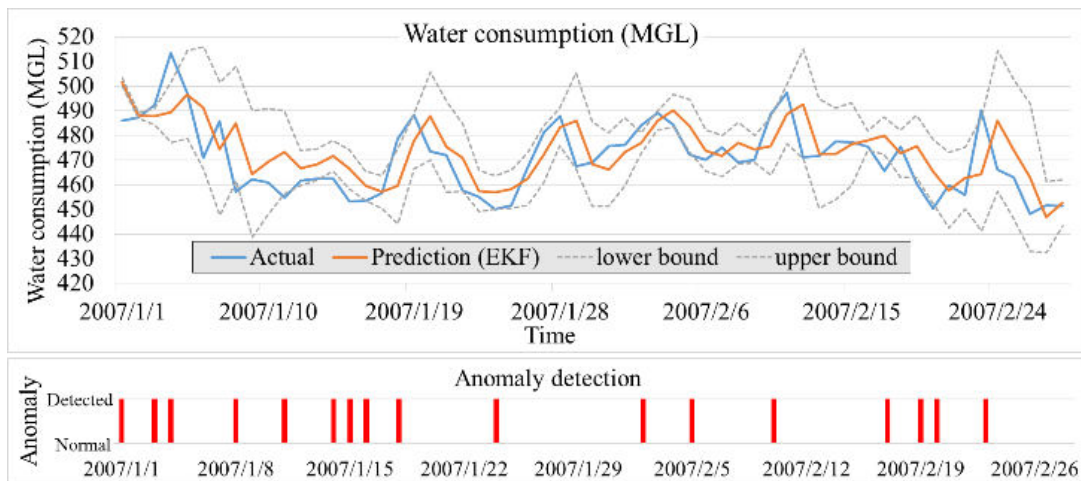
Figure 4 Anomaly detection for daily water consumption

As briefly elaborated earlier, the confidence interval of the prediction can be calculated by EKF, and this information can be used to detect the anomaly in the system. In this case, by running the EKF while the DNN is predicting, the upper and lower bounds of the prediction can be obtained. By choosing a 2σ confidence interval that σ is standard deviation, the actual water consumption will have about 95% possibility to be within the bounds. If the actual measurement is out of the bounds, it indicates that there is a very high chance that something wrong happened in the system. Then, a warning could be issued to inform system operators. Figure 4 demonstrates the examples of the anomaly detection for this case.

3.2 15-Minute water consumption prediction

The second case study is also the water consumption prediction problem of the Oldham town, but the data is recorded in every 15 minutes. The model is built to input the previous 2.5 hours' data to the neural network and predict the next 15-minute water consumption. The deep belief neural network model has been optimized with 10 inputs and 1 output, since the data have 15 minutes' interval. Meanwhile, the neural network has 3 hidden layers with 7 neurons in each layer.
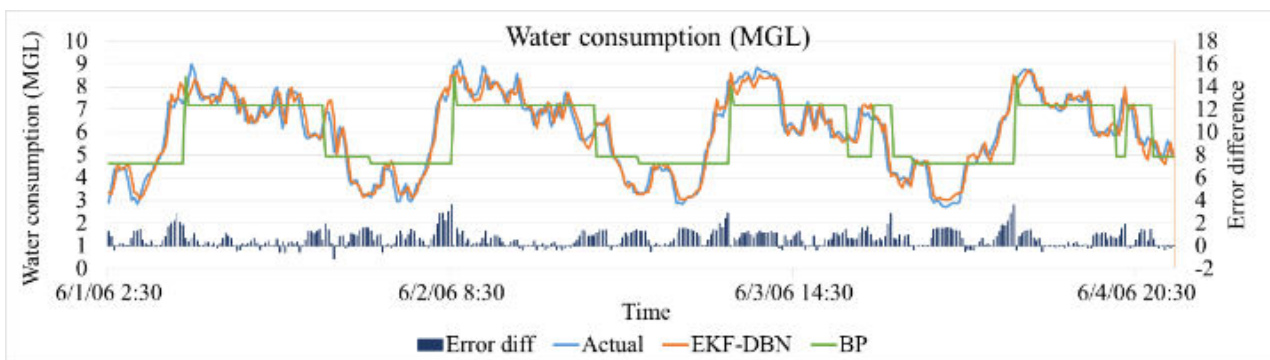

Figure 5 Comparison of predictions and the actual 15-min water consumptions

As shown in Figure 5, the prediction from EKF-DNN model can track the actual water consumption very well. The MAPE of the prediction is 5.543% and the RMSE is 0.2. Therefore, it can be concluded that the EKF-DNN model is very effective to achieve adequately accurate prediction of water consumptions and anomaly detections. The result of the proposed method is compared with the BP method. It shows that the proposed method outperforms the BP method in this case study. This is because the data has a period of one day, but the 2.5 hours input data is not enough for the neural network to learn the pattern. BP method without function of data assimilation, the prediction result is less accurate than EKF-DNN.

4   CONCLUSIONS

This paper presented a generic data analytic modelling framework for extracting the intelligence from data, with potential of wide applications in simulation and prediction analysis that serves as the data-driven engine of infrastructure asset performance modelling e.g. water distribution network management. The results of the case studies conducted and presented in the paper demonstrate the effectiveness of the integrated approach for practical applications.

References

[1]   Z. Y. Wu and A. Rahman. "Optimized Deep Learning Framework for Water Distribution Data-Driven Modeling." *Procedia Engineering*, 186, 261–268, 2017.

[2]   G. E. Hinton, S. Osindero and Y.-W. Teh "A fast learning algorithm for deep belief nets." *Neural computation*, 18(7), 1527-1554, 2006.

[3]   G.A. Einicke and L.B. White. "Robust Extended Kalman Filtering." IEEE Trans. Signal Processing. 47 (9): 2596–2599. *doi:10.1109/78.782219*, 1999.

[4]   Z. Y. Wu, Q. Wang, S. Butala, T. Mi and Y. Song. "Darwin Optimization Framework User Manual." *Bentley Systems, Incorporated, Watertown, CT*, USA, 2012

[5]   V. Nair and G. E. Hinton, Rectified linear units improve restricted Boltzmann machines. *Proc., Proceedings of the 27th international conference on machine learning (ICML-10)*, 807-814. 2010.

[6]   Z. Y. Wu and X. G. Yan. "Gene Expression Approach for Short-Term Water Demand Forecasting", WDSA2010, Sept 11 – 13, 2010, Tucson, Arizona, USA.