Hybrid Neuron Network and Random Forest Model for Smart Meter Abnormal Detection

Mengmeng Zhuang, Zuyi Li, Senior Member, IEEE, and Mohammad Shahidehpour, Fellow, IEEE,

*Abstract*—**Smart meter abnormal detection is very important for smart meter maintenance and fire prevention, electricity anti-stealing and so on. Machine learning is a promising method that can learn hidden information and more exact features of smart meter data. We adopted two different common approaches using Artificial Neural Network (ANN) with Random Forest (RF) to improve detection accuracy. The proposed approach was implemented to estimate the abnormal changing of smart meter. The investigations provide an effective reference for solving smart meter abnormal detection by taking smart meter hot socket as an example.**

***Index Terms*—Machine learning; Smart meter abnormal detection; Artificial Neural Network; Random Forest**

# Introduction

W

ITH the deployment of advanced metering infrastructure (AMI), with around 500 million deployment expected to install smart meter by 2020 [1], are considered as the first step of moving into smart grids for utilities. With smart meter, utilities can allow for monitoring power usage by remote control [2], better understand the rational design of electricity demand response systems [3], short-term load forecasting system [4], the behavior of electricity customers [5], energy theft and cyber security [6-7], outage management [8] and so on. However, smart meter fires reported in recent years have posed a threaten to public concern, such as explosions, electrical appliance burned out. From 2011 to 2013, smart meters were linked to 23 incidents reported to Ontario’s Fire Marshal in Canada [9]. In 2012, Electrical Trades Union Australia called for suspending the installation of smart meter because of the fire accidents [10]. In Canada, the government both in Saskatchewan and Portland Oregon decided to replace 175,000 smart meters with traditional meters because of several fires [11]. Lakeland Electric replaced 10,657 Residential Smart Meters because of six residential meter fires [12]. Smart meter caused $75,000 in damage to residence because of exposition in Virginia State in the USA in January 7, 2015[13] - [14]. Such accidents caused by smart meter caused the deployment of smart meter face great challenge. It is necessary and urgent to enhance the AMI management and detect abnormal smart meter in advance, analyze the cause of smart fires and improve smart meter maintenance efficiency.

A variety of causes can contribute to the smart fire incidence, such as improper installation, aging, change between metal and ceramic parts to plastics, and so on. All these causes can be explained as hot socket phenomenon. Hot socket is a very common problem that cause fire incidence and has been for decades. In literature [2], some experiments have confirmed the hot socket phenomenon caused by some the arousing of smart meter temperature internally, the current between two poles of meter and ambient temperature. The varying of temperature is the cause of any fires. Similarly, smart meter temperature forecast, and monitoring is very crucial in preventing smart meter fires. However, smart meter temperature varying is a continuous, data-intensive, dynamic and chaotic process and relevant with some other unexpected factors [2]. In our paper, we will firstly analyze the factors that cause the rising of smart meter temperatures and model the temperature varying of smart meter to further explore the cause of smart meter fires based on the study in [2].The parameters required to estimate smart meter temperature changing are enormously complex so that there is uncertainty in prediction even for a short period. Conventional approaches, such as time-series [15], autoregressive model [16], fuzzy control [17], grey prediction [18] and so on, fail to accurately represent the complex un-linear relationship between the varying temperature of smart meter and some uncertainty factors.

Machine learning, as a powerful tool, artificial neural networks, such as have received a great deal of attention to solve some forecasting problems [19]. In the field of electricity, ANN was successfully used to predict short-term wind power in Portugal [20], short-mid-term solar power [21], short load forecast [22], electricity price forecast [23] and annual electricity consumption forecasting [24]. A common and simple approach often adopts back-propagation (BP) algorithm [25]. In recent years, decision tree, another practical, easy to understand and simple machine learning approach, has attracted great attention among research study [26]. Random Forest (RF) is the most popular class, which can overcome conventional disadvantages of decision trees[[27]](https://www.sciencedirect.com/science/article/pii/S0378778816313937#bib0150) , and enable ensemble learning via voting scheme [[28]](https://www.sciencedirect.com/science/article/pii/S0378778816313937#bib0155).

Our contribution in this paper:

* Hot socket phenomenon is considered as one example to study smart meter abnormal problem.
* The most important features of raw smart meter data are analyzed statistically.
* Two different approaches were adopted for improving accuracy of identifying abnormal smart meter.

The rest of this paper is organized as follows: Two different approach is proposed to solve smart meter abnormal detection in Section II; Case studies are verified the proposed approach using real dataset and system model in Section III; Finally, we draw some remarks conclusions in Section V.

# Methodology

## Hot Socket Phenomenon

Previous study only study smart meter data related factors, however, few studies do the study of smart meter temperature for detecting abnormal smart meter or not. The factors that caused hot socket have been in detailed in [2], which can be seen in Fig.1.

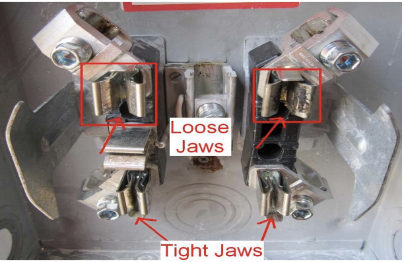


Fig. 1. Hot socket of smart meter [2]

According to [2], A mathematical relationship between the total dissipated power at the meter base and the temperature sensor readings was found, correlating the expected temperature reading of the sensor to the applied current through each of the two poles and the ambient temperature, which can be formulated as

 (1)

Where is active aggregated appliances number, is one time period, and represent the current through each of two poles of smart meter. The experiments in [2] are conducted in ideal environments. In the practical applications, more factors should be considered.

From the analyzing above, we find the smart meter temperature change have roughly linear relationship with weather temperature, but no specific relationship with load data. So, we decide to adopt to build the non-linear model of smart meter in order to improve the forecasting accuracy.

## Artificial Neuron Network

The development of ANN model is based on studying the relationship between input variables and output variables. Basically, the neural neuron architecture consists of three or more layers, i.e., input layer, output layer and hidden layer as shown in Fig.2.The model of neural networks can be descripted as:

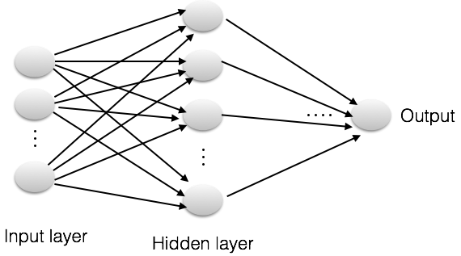


Fig.2. The Structure of Artificial Neuron Network

The neuron can receive the neuron at time ,

input , the output of neuron  is :

 (2) Where  is the synapsis connecting parameter, namely, weights,  is the threshold value of neuron ;  is the synapsis time extension between input, output;  is the activate function of neuron network.

Our proposed model can be summarized as follow in Fig.3. It was observed from the half hourly load data of half one year for Comed Inc. The meter data obtained from the GE I-210. The study data reported in this paper, uses hourly data by the method of disposal data, such as interpolation. The paper uses hourly historical meter temperature data, load data and weather information for an area in the state of Illionis collected by Comed Inc and Weather Source Station, Chicago, IL.

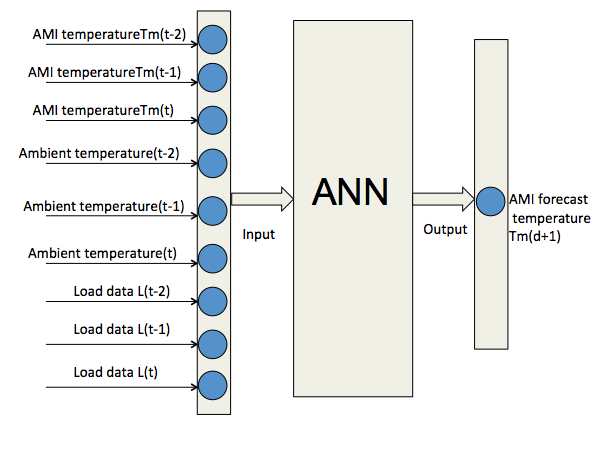


Fig.3. Proposed ANN model structure for abnormal detection

## Random Forest

Random forest approach is made up of a series of decision tree, whose structure is made up of random vector. If decision tree is considered as one expert of task classification, then FR is one classification of many experts focus on task. In this part, we build classification and prediction model using random forest (RF) algorithm with a hybrid feature. Here we divide abnormal or normal smart meter into positive sample and negative sample.



Fig.4. Random forest model structure

We choose the eigenvector of the starting point at each moment of smart meter feature data. And the features include

smart meter ID, location, meter temperature, weather temperature, and load data.

# Case Study

For validation, the proposed model approaches are coded in MATLAB and Python 3.7. All tests are based on the DELL XPS8500 Desktop with key specs: Intel(R) Core (TM) i7-3770 CPU @ 3.40GHz, 4,096.0MB x 2, Microsoft Windows 10 Pro.

## Data Preprocessing

The data of our forecasting proposed model is 200 meters data provided by ComEd.inc, a utility company. We will analyze raw data of smart meters provided by ComEd company and extract features of data. Collected smart meters data information include meter numbers, 3 models (GEI120+, I120+C, Focus), the location of installations (17 types), towns, load (48 half-hourly), and weather temperature in Illinois state (24 hourly), time (one year). Besides, we download the ambient temperature from NOOA website: <http://www.ncdc.noaa.gov/>. Because the weather condition of the city nearby Chicago is similar with that of Chicago, we consider all the cities have the same temperature.

The key point of data mining is to clean the raw bad data, redundant data and transform it as what we needed. Before experiments, we conduct data processing. Data cleaning and disposal are very crucial in machine learning. Firstly, data cleaning includes missing data and noisy data. For some missing data, we adopt fetching mean value between two value. Wavelet transform [29] method was adopted to disposal some volatility or bad data. Secondly, data transformation include normalization, attribute selection, discretization and concept hierarchy generation. Also, data reduction include data cube, aggregation, attribute subset, selection, numerosity reduction and dimensionality, or some redundant information, primary component analysis (PCA) [30] approach was adopted to reduce the dimension of features and extract valuable information from massive data.

## Smart Meter Temperature Data Analysis

Smart meter temperature changing is the most key factor feature to identify if a smart meter is abnormal or not. So, we analyzed the smart meter temperature data statistically. In order to better show the temperature distribution of smart meter, we investigate the Gaussian Distribution of smart meter temperature, as can be seen in Fig.5. From Fig.5, we can find  is 72.7367, is 25.9023. Also, we deeply analysis smart meter temperature distribution characteristics in different hours, week, and month, which were illustrated respectively in Fig.6, Fig.7 and Fig.8. From these figures, wecan find that the distribution of smart meter temperature data is approximately normal distribution.

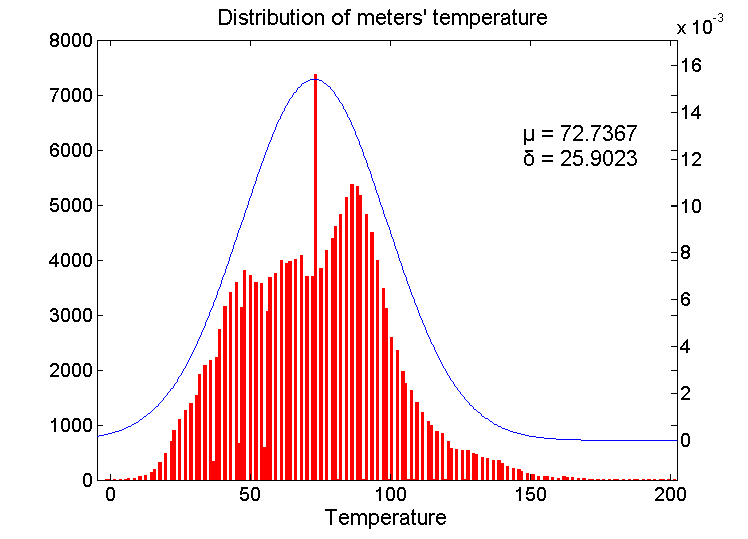


Fig.5. Gaussian distribution of smart meter temperature

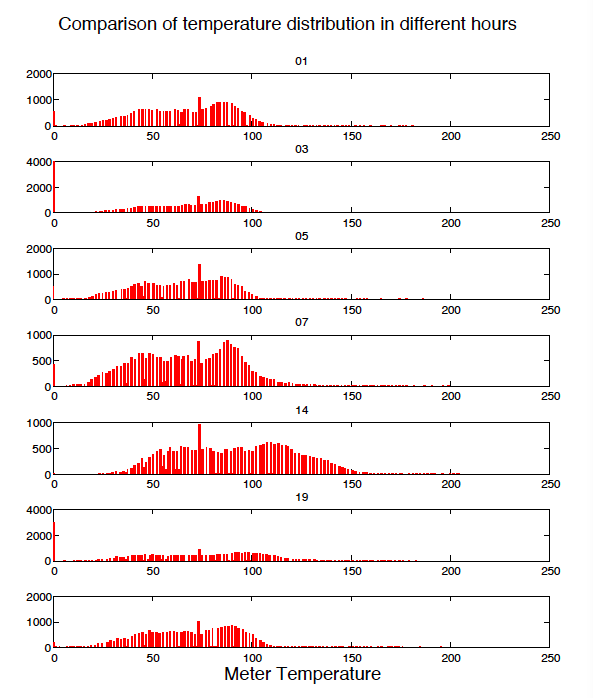


Fig. 6. Comparison of smart meter temperature distribution in different hours

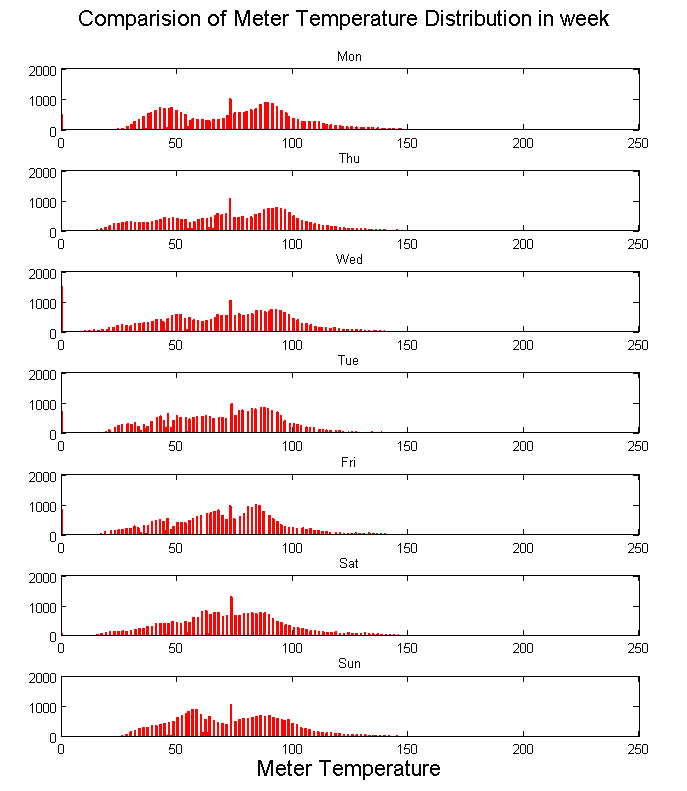


Fig.7 Comparison of smart meter temperature distribution

in weeks

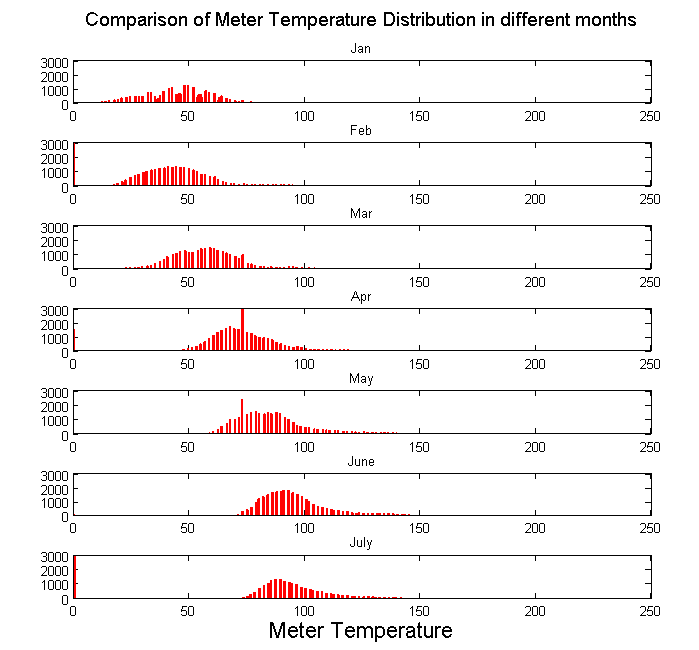


Fig.8. Comparison of smart meter temperature in months

## Matrix

The measure of error between the actual and predicted meter temperature is obtained using average percentage error (APE) and the mean absolute percentage error (MAPE).

 (3)

 (4)

Where is the smart meter forecasted data,  is the smart meter actual data, is the number of hours in the forecasting period.

## Results and Discussions

As for the aspect of abnormal detection of smart meter, the applications of ANN and RF can be summarized as two aspects: one is to build forecast model to predicate the trend on the side of forecast of smart meter data by adopting the proposed model ; another one is to apply the proposed model as an classifier to diagnose and reason from the aspect of pattern recognition.

### ANN results

The input data for ANNs are normalized such that they fall within the range [-1,1]. The training models are with 12 neurons for the hidden layer, and 1 neuron for an output layer. From the above analyzing, we can see that the hidden layer number is actually related with the number of the input number. For the same issue, the simper the model structure is, the stabber network model is. So, when choosing the network model, we should first consider the simple structure as much as possible. The data covers from Jan.2014 to May.2014 whereas model testing is investigated by using year 2013, to test the generalized error forecasting of the ANN. One external variable to be used as input so that one input to network further 1 neuron are used for hidden layer. In this project, we define 9 inputs, the smart meter temperature data Tm at the day before yesterday-Tm (t-1), the load data at the day before yesterday-L(t-1),The ambient temperature data-Tw(t-1), load data at tomorrow-L(t) and ambient temperature data at temperature-Tw(t).There must be one neuron in the output layer because there is only one target value associated with each input. The output is the smart meter temperature tomorrow forecast Tm(t). Back-propagation (BP) algorithm [25] was adopted in the proposed model.

From the Fig.9, we can find the network has a good performance in the training of ANN; the epoch is 3822 when the results meet the desired results. From figure.6, the ANN model has a better forecasting fitting result than the conventional WMLR approach. The MAPE of two approaches –WMLR and ANN is respectively 7.1681% and 2.3193%. The biggest APE of WMLR approach is up to 14.5269%, while adopting ANN approach is up to 6.4395%.

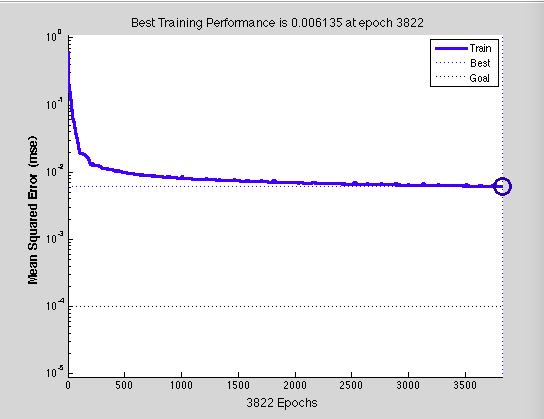


Fig.9. Training performance of ANN

TABLE I

ANN Approach Results

|  |  |  |  |
| --- | --- | --- | --- |
| Hour | WMLR (%) | BP-NN (%) | |
| 0  1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23 | 6.1049  2.7175  0.3818  3.1245  5.0620  8.5605  7.5911  8.3193  7.8030  5.6231  2.6570  1.2588  3.1788  4.9333  7.5757  4.9796  7.7952  13.3774  14.5629  13.5836  11.6917  12.2529  10.1164  8.7839 | 2.3511  2.8400  3.6965  4.2624  4.5889  0.8703  0.1887  2.4428  1.8622  1.1238  1.6250  2.8180  4.5030  6.4395  1.9262  2.0899  0.6615  0.0189  0.0604  1.8972  1.2567  6.3439  0.8809  0.9149 |
| MAPE (%) | 7.1681 | 2.3193 |

### Random forest for classfication forecasting

Feature choosing strategy is very important. In order to improve our model accuracy, we can have different feature strategy. First one is to use temperature and load as feature, training set is the past T time period. Dataset format for training set: [Mt-1, Wt-1, It-1, Mt-2, Wt-2, It-2, …Mt-T, Wt-T, It-T].Dataset format for testing set: [Mt] [Wt., It], and T can be determined four points of half day,7 points of one day, 210 points of one month. Second one is to choose 21 attribute value as feature before 7 points of t periods. And choose the case of during T period as training set. In order to identify abnormal smart meter effectively, we divide temperature into three intervals [0-100], [100-120], and [120, +∞] as normal smart meter, approximate normal smart meter and abnormal smart meter. And we get the experiment results, as illustrated in TABLE II.

TABLE II

RF Approach Results

|  |  |  |
| --- | --- | --- |
| Category | Error | Forecast /Actual Points |
| Whole dataset | 5.0718% | 165100/173921 |
| Tm<100 | 0.2372% | 149704/150060 |
| 100<Tm<120 | 31.807% | 11350/16644 |
| 120<Tm | 43.937% | 4046/7217 |

Tm is the temperature of smart meter

### Results discussions

We compare two different approaches and find first forecasting result is better than second one. Compared with ANN approach, RF approach forecasting accuracy is a little lower than ANN approaches. However, for disposal and model forecasting speed are very fast, can overcome the drawbacks of overfitting of ANN approach, especially for large scale of implementation.

TABLE III

Comparations Between ANN and RF Approaches

|  |  |  |
| --- | --- | --- |
| Approach | Detection/Classification Accuracy | Model  Performance |
| ANN | Relative High/Low | Time consuming, easy to overfit |
| RF | Relative Lower/High | Fast, not easy to fall into overfit |

# Conclusions

This paper proposed hybrid ANN-RF model for smart meter abnormal detection. Smart meter temperature is the most important feature to detect abnormal smart meter. The forecasted models have enhanced accuracy due to applying ANN approach with BP algorithm to the network model compared with conventional linear regression approach. Our results also show that ANN show good performance for forecasting while RF show good performance for classification with very fast speed without overfitting. It is suggested that in the practical implementation, a hybrid ANN-RF model should be adopted to improve abnormal detection accuracy by integrating the advantages of ANN for perception and RF for classfication with fast speed. This paper introduces for a reference to enhance the smart meter capability of abnormal detection.

# Acknowledgement

The data in this study is based on the data center of Comed.Inc. I would like to acknowledge with gratitude, the support of the ComEd Inc. The author wishes to take this opportunity to express sincere appreciation for the support from ComEd, Inc.

# Reference

[1] D.G Hart, “Using AMI to realize the smart grid,” Proc.*IEEE* Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy,Pittsburgh,PA, July 2008, pp. 1–2.

[2] Paul W.Brazis, Jr. Evaluation of Smart Meter.

[3] Temperature Sensing for Remote Socket Overheat. September 25,

2014.

[4] Conejo A J，Morales J M，Baringo L．Real-time demand response model[J]．IEEE Transactions on Smart Grid，2010，1(3)：236-242.

[5] Franklin L. Quilumba, Wei-Jen Lee, Fellow, Heng Huang, David Y. Wang, and Robert L. Szabados, Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities [J] *IEEE* Transactions on Smart 2015,2(6): 911-918.

[6] K. Tweed, “Fraud should be first priority for smart meter security,” Available:https://www.greentechmedia.com/articles/read/fraud-should-be-first-priority-for-smart-meter-security/.

[7] F.M. Cleveland, “Cyber security issues for advanced metering infrastructure,” Proc. *IEEE* Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy, Pittsburgh, PA, July 2008, pp. 1–5

[8] “Business agility in outage management for energy and utilities,” IBM Software Energy Utilities. 2012.

[9] “Smart meters linked to 13 fires in Ontario, Fire Marshal says”, accessed on: http://www.thestar.com/news/canada/2014/08/08/

smart\_meters\_linked\_to\_13\_fires\_in\_ontario\_fire\_marshal\_says.html.

[10] Electrical Trades Union again calls for suspension of smart meter rollout <http://www.heraldsun.com.au/archive/news/smart-meter-death-fears/story-fn7x8me2-1226285463342>.

[11] “Accidental Alexandria Area House Fire”,

<http://www.fairfaxcounty.gov/fr/news/2015archive/2015_04.htm>.

[12] “Fires prompt removal of 175,000 smart meters in Canada andOregon”.http://emfsafetynetwork.org/fires-prompt-removal-of-175000-smart-meters-in-canada-and-oregon/.

[13] “Overheating Concerns: Lakeland Electric to Replace 10,657 Residential Smart Meters” <http://www.theledger.com/article/20140826/NEWS/140829388/1410?Title=Overheating-Concerns-Lakeland-Electric-to-Replace-10-657-Residential-Smart-Meters->

[14] January 7, 2015:  Alexandria, Virginia: Smart meter explodes and causes $75,000 in damage to residence.[http://www.fairfaxcounty.gov/fr/news/2015archive/2015\_04.htm#](http://www.fairfaxcounty.gov/fr/news/2015archive/2015_04.htm)

[15] Y. Chen, B. Yang, Q. Meng, Y. Zhao, and A. Abraham. Time-series forecasting using a system of ordinary differential equations. Information Sciences, 181(1):106–114, 2011.

[16] Kristiansen T (2012) Forecasting Nord Pool day-ahead prices with an autoregressive model. Energy Policy 49:328–332

[17] S. Sachdeva and C. Verma, "Load forecasting using fuzzy methods," in Joint International Conference on Power System Technology and IEEE Power India Conference, pp. 1-4, October 2008.

[18] P. Zhou, B.W. Ang, K.L. PohA trigonometric grey prediction approach to forecasting electricity demand Energy, 31 (2006), pp. 2839–2847

[19] W. L. Gorr, Research prospective on neural network forecasting, Int. J. Forecast; 1994. vol. 10, p. 1–4.

[20] Catalão, J P S, Pousinho, HMI and Mendes, VMF (2009). An Artificial Neural Network Approach for Short-Term Wind Power Forecasting in Portugal. Eng. Intell. Syst. Elect. Eng. Commun,17(1) 5–11.

[21] E. Izgi, A. Öztopal, B. Yerli, M.K. Kaymak, and A.D. Sahin, “Short-mid-term solar power prediction by using artificial neural networks”, Solar Energy, vol. 86, no. 2, pp. 725–733, 2012.

[22] Changhao Xia; Zhonghua Yang; Hongjie Li, "Electric load forecasting using virtual instrument based on dynamic recurrent Elman neural network," Power Engineering and Automation Conference (PEAM), 2012 IEEE, vol., no., pp.1,4, 18-20 Sept. 2012.

[23] Peter S E, Raglend I J and Simon S P “An Architectural Frame Work of ANN Based Short Term Electricity Price Forecast Engine for Indian Energy Exchange Using Similar Day Approach”, IMPACT: International Journal on Research & Engineering Technology, (2014), Vol. 2, No. 4, pp. 111-122.

[24] Azadeh, A., Ghaderi, S.F., Sohrabkhani, S. Annual Electricity Consumption Forecasting By Neural Network In High Energy Consuming Industrial Sectors.[J]. Energy Conversion and Management, 2008 49(8):2272–2278.

[25] Yangming, Guo & Congbao, Ran & Xinyu, Ji. 2013. Analogous circuit fault diagnosis based on combinatorial optimization BP neural network. Northwestern Polytechnical University Journal. 31(1): 45-48.

[26] B. Larivière, D.V. den Poel Predicting customer retention and profitability by using random forests and regression forests techniques Expert Syst. Appl., 29 (2) (2005), pp. 472- 484, [10.1016/j.eswa.2005.04.043](https://doi.org/10.1016/j.eswa.2005.04.043) ISSN 0957-4174.

[27] L. BreimanRandom forests Mach. Learn., 45 (1) (2001), pp. 5-32

[28] R. Jiang, W. Tang, X. Wu, W. Fu. A random forest approach to the detection of epistatic interactions in case–control studies,BMC Bioinform., 10 (1) (2009), p. 1.

[29] C. K. Chui, An Introduction to Wavelets. New York: Academic, 1992, pp. 6–18.

[30] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," Automatic Control, IEEE Transactions on, vol. 26, pp. 17-32, 1981.

**Mengmeng Zhuang** is currently a Research Associate in the Robert W. Galvin Center for Electricity Innovation at Illinois Institute of Technology, Chicago, IL, USA.

**Zuyi Li (SM’09)** is currently a Professor and the associate director of the Robert W. Galvin Center for Electricity Innovation, Illinois Institute of Technology, Chicago, IL, USA.

**Mohammad Shahidehpour (F’01)** is the Bodine Chair Professor and the director of the Robert W. Galvin Center for Electricity Innovation, Illinois Institute of Technology, Chicago, IL, USA.