Learning-based Data Analytics: Moving Towards Transparent Power Grids

Kunjin Chen, *Student Member*, *IEEE*, Ziyu He, Shan X. Wang, *Fellow*, *IEEE*, Jun Hu, *Member*, *IEEE*, Licheng Li, and Jinliang He, *Fellow*, *IEEE*

***Abstract*—In this paper, we present the learning-based data analytics moving towards transparent power grids and provide some possible extensions including machine learning, big data analytics, and knowledge transferring. The closed loops of data and knowledge are illustrated and the challenges for establishing the closed loops are discussed. General ideas and recent developments in supervised learning, unsupervised learning, and reinforcement learning are presented together with extensions for power system applications. Furthermore, much emphasis is placed on privacy-preserving data analysis, transfer of knowledge, machine learning for causal inference, scalability and flexibility of data analytics, and efficiency and reliability of computation. Existing integrated solutions in the industry featuring the Industrial Internet and the digital grid are also introduced.**

***Index Terms*—Data analytics, machine learning, smart grid, transparent power grid.**

I. Introduction

THANKS to the increasing availability of large amounts of data in the power systems, it is expected that data analytics will play an increasingly important role in future smart grids [1]. In addition to traditional measurement devices, various types of data collection units including remote terminal units (RTU), phasor measurement units (PMU), and smart meters are being widely deployed all over the world. With the development of sensor technology, connecting large numbers of sensors into wireless sensor networks (WSN) has also become a trend in various industries in order to capture and create value from data generated by machines, devices, as well as people [2].

The scope of modern power systems has greatly expanded with the integration of distributed generation (DG), multi-energy system (MES) and active distribution network (ADN) [3], [4]. The massive amounts of data collected by low-cost and high-quality sensors are critical to the effective monitoring, control, and analysis of the various components of smart grids. At the same time, the rapidly increasing volume of data, the heterogeneousness of data collected by different types of sensors, the physical constraints of the power grids, and the diverse nature of various applications within the power and energy systems all pose great challenges for data analytics and knowledge discovery. In this paper, the term "data analytics" refers to the process of examining, cleaning, transforming, and modeling data in order to draw conclusions about the information contained in the data and to support decision-making, and the term "knowledge discovery" refers to the process of discovering useful knowledge with high generality. The discovered knowledge can be very abstract so that it can be effectively shared among relevant entities.

From the perspective of sustainable development, clean energy and low-carbon growth are among the core topics of power and energy industries. For instance, the European Union planned in 2014 to increase the proportion of renewable energy to 27% by 2030 [5]. China also has plans to increase the proportion of non-fossil fuels in primary energy consumption to around 20% by 2030 [6]. In order to enhance environmental friendliness and sustainability along the whole power and energy industry chain (production, transmission, distribution and consumption), well-directed and comprehensive data collection is required. System-level optimization based on collected data is able to increase the life-cycle efficiency of energy and reduce emissions. In addition, environmental regulatory agencies and carbon markets can also benefit from the information provided by the data collection infrastructure.

With the help of low-cost and high-quality measurement devices and sensors networks, the detailed real-time information of any interested nodes in the power grids can be collected, which means that the power grids are transparent and can be called "transparent power grids." The concept of transparent power grids, which is not an alternative but a further modification to the concept of smart grids, is highlighted in this paper in order to show how proper data analytics and learning methods can jointly make the power grids more observable, analyzable, predictable, and controllable along with traditional methods. In Fig. 1, the features of transparent power grids are illustrated. The development of the data acquisition infrastructure lays the foundation for the all-round monitoring of power grids. The

Manuscript received October 12, 2017; revised January 12, 2018, accepted January 29, 2018. Date of publication March 30, 2018; date of current version January 28, 2018.

K. Chen, J. Hu, and J. L. He (corresponding author, e-mail: hejl@tsinghua.edu.cn) are with the State Key Lab of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China.

Z. He is with the Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90007, USA.

S. X. Wang is with the Center for Magnetic Nanotechnology, Stanford University, Stanford, CA 94305 USA, and also with the State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China.

L. C. Li is with the Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China, and China Southern Power Grid, Guangzhou 510623, China.

DOI: 10.17775/CSEEJPES.2017.01070

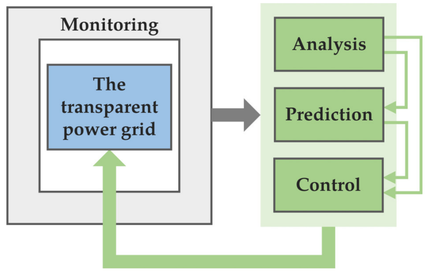


Fig. 1. The features of transparent power grids.

abilities of analysis, prediction, and control are equally important for establishing power grids with high transparency. In order to build a transparent power grid, a robust and efficient big data system connecting data sources applications is needed [7]–[9]. In addition to the general ideas of big data analytics, the concept of learning from data is also gaining widespread attention. While well-established machine learning techniques have already been widely adopted in a variety of applications in power systems, new models and methods have been developed and implemented in many fields over the last few years. Thus, it is of great importance to survey the recent advances in learning-based data analytics and explore some potential applications in power systems, especially for the aspects pointed out in Fig. 1.

In this paper, we give a comprehensive in-depth survey of learning-based data analytics in the context of transparent power grids. The remainder of this paper is organized as follows. In Section II, we introduce the features of transparent power grids. The closed loops of data and knowledge are also discussed in detail. The challenges for establishing the closed loops of data and knowledge are presented in section III. Integrated solution examples are briefly mentioned in section IV. Finally, the paper is concluded in Section V.

II. The Closed Loops of Data and Knowledge Moving Towards Transparent Power Grids

In this section, we first briefly introduce the concept of transparent power grids, followed by the demonstration of the closed loops of data and knowledge. Then, we survey and extend the learning methodologies that use data to discover knowledge in transparent power grids.

A. The Features of Transparent Power Grids

We elaborate the concept of transparent power grids through the four features mentioned in Fig. 1 as follows:

1) *Monitoring:* One of the major features of the transparent power grid is that it monitors the events occurring in the power system with great detail and accuracy, so that all the key information in the power system will not be overlooked. In today's power grids, however, the data acquisition infrastructure is not mature enough and has great room for development. Currently, different types of data collecting devices including RTU, PMU, and smart meters are being widely deployed in the power grids. Wide area monitoring systems (WAMS) and advanced metering infrastructure (AMI) are also being established. In addition, with the development of low-power-consumption and low-cost sensors, the application prospect for WSNs in the power system is also very broad. Generally speaking, in a transparent power grid, we will have broader data sources, and the data we collect will have diverse types and structures.

2) *Analysis:* The ability to analyze collected data is the core functionality of a transparent power grid, which allows full utilization of data continuously generated from all corners of the power grid. At the same time, it is also the field where various data analysis methods and machine learning methods are directly applied. Compared with data analysis in a traditional power grid, the analysis in a transparent power grid is more dependent on the power of the data and our ability to obtain knowledge from the data. The interactions among data, knowledge, and human will also be the keys to the further development of power system analysis.

3) *Prediction:* A critical aspect in the development of artificial intelligence is to allow the machines to have the ability to predict, which can also bring higher transparency to the power systems. Not only do we need to grasp what is happening in the past and the present, but we also need to know more about what could happen in the future with greater confidence and capture the uncertainties of the possible outcomes. The ability to predict relies heavily on the development of data analytics and machine learning techniques and also on our understanding of the physical models and underlying mechanisms of the power systems. Theoretical knowledge related to the dynamics of power systems should also be accumulated so that we can simulate the power systems with high precision.

4) *Control:* Being able to control means we cannot only analyze the data and predict the future, but also interact with the power systems in a timely and efficient manner. In contrast to the control requirements in a traditional power grid, the control in a transparent power grid should be based on the organic combination of the abilities to monitor, analyze, and predict, so that the short-term and long-term results of the control decisions can be taken into full consideration. The development of reinforcement learning (RL) provides an idea for control in a transparent grid, which will be elaborated later in this paper. At the same time, traditional control methods could also benefit from the above-mentioned features of transparent power grids.

Establishing closed loops of data and knowledge is of great importance for the monitoring, analysis, prediction, and control in future power systems. In the next subsection, we will introduce the closed loops of data and knowledge as well as the machine learning techniques related to the closed loops.

B. The Overview of the Closed Loops of Data and Knowledge

The volume of data within the power grids, as previously mentioned, is increasing at a rapid pace. In addition, one of the key purposes of learning-based data analytics is to discover or acquire knowledge that can be explicitly expressed (or understood) and shared. We expand the data analytics and knowledge discovery framework presented in existing studies [10]–[12] by introducing the closed loops of data and knowledge. An illustration of the data and knowledge closed loops in smart grids and transparent power grids is shown in Fig. 2. In addition to the flow of data from data sources to

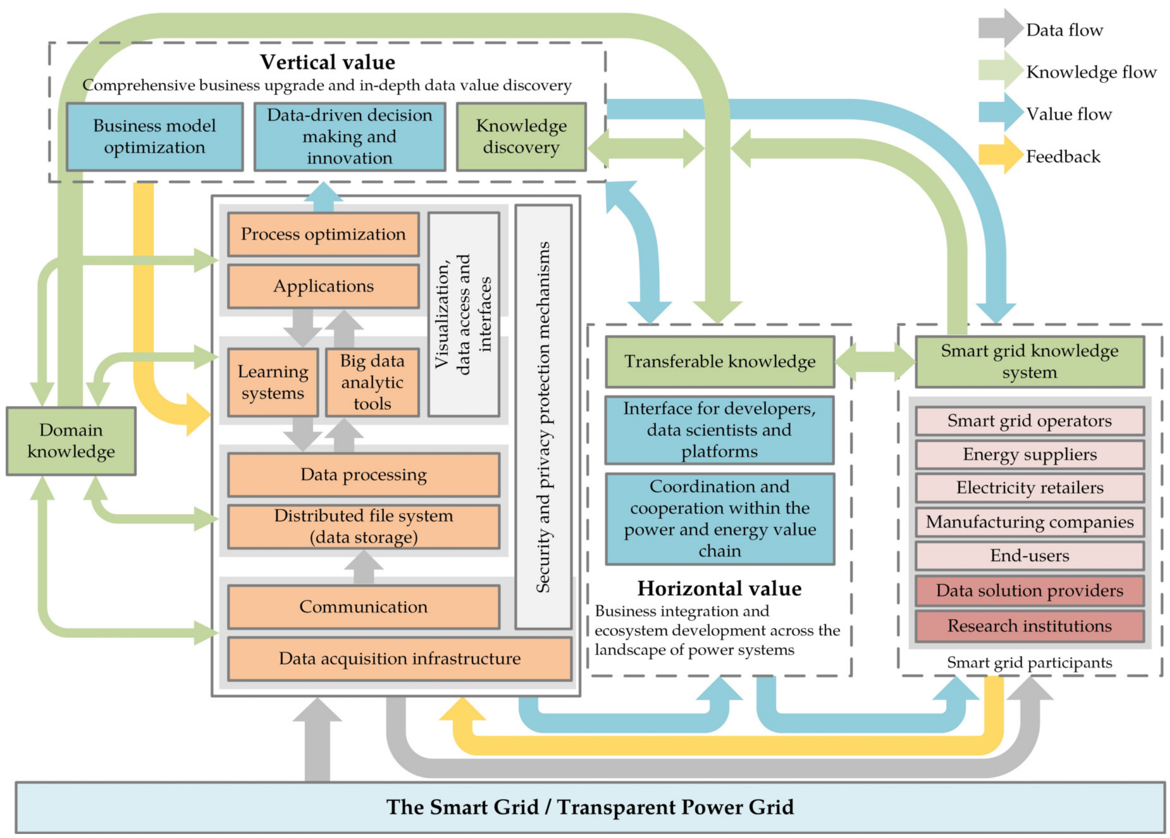


Fig. 2. The closed loops of data and knowledge towards transparent power grids. The green arrows denote the flow of knowledge and the grey arrows denote the flow of data. Value flows are denoted using blue arrows. Yellow arrows are used to represent feedbacks.

applications within the data analytics framework, the security and privacy protection mechanisms are given high priority, as they are involved in almost all aspects where data exists [1]. The vertical and the horizontal values are attached to the data analytics framework in two directions. Another block containing the smart grid participants also appears in the figure in order to emphasize that humans are included in the closed loops. The detailed descriptions of the arrows are as follows:

1) *The data flows:* The data generated in the power systems are first collected by the data acquisition infrastructure. Then, the data flows through the entire data processing and analysis procedures. The Closed loops of data in vertical markets can be formed if data generated within applications and businesses related to the data analytics framework can be collected. Although direct and unlimited data sharing across horizontal markets is impractical, the sharing of desensitized datasets among smart grid participants is possible. A well-established data sharing mechanism enables smart grid researchers, data scientists, and data solution providers to use the data more efficiently and effectively, thus contributing more to the smart grid knowledge system.

2) *The knowledge flows:* First, domain knowledge can be utilized in the procedures in the data analytics framework. Newly discovered knowledge and the knowledge system of smart grid participants can interact with existing domain knowledge. By integrating and comparing knowledge from different sources, transferable knowledge can be identified and then added to the smart grid knowledge system shared by smart grid participants. Thus, the closed loops of knowledge can be established. Note that the different blocks of knowledge are not mutually exclusive.

3) *The value flows:* The values provided by the data analytics framework can propagate both vertically and horizontally, benefiting the smart grid participants. Furthermore, the vertical value and the horizontal value can be coordinated, as companies may own businesses in different vertical markets that are along the same value chain. Cooperation among smart grid participants can also enhance the connection between the vertical value and the horizontal value.

4) *The feedbacks:* The first feedback to the data analytics framework comes from the vertical value block, which is important for the iterative development of the procedures within the framework. The second feedback comes from the smart grid participants who are related to the data analytics services.

More specifically, we extend the framework in both vertical and horizontal directions. In the vertical direction, we primarily focus on the business model and the decision making process related to a specific application or applications that are similar in nature. With the help of in-depth data analytics and knowledge discovery, the business models can be optimized, and more decisions can be made in a data-driven manner. Both the optimization procedure and innovations can reach down and improve the functionality of the entire framework.

Horizontally speaking, the framework can be extended in two aspects. First, interfaces need to be open to application developers, data scientists, and platforms that provide data analytics and learning solutions. This could greatly benefit the development of a community dedicated to providing coherent and effective solutions to the storage, processing, and mining of data, bringing greater value to the power and energy systems in the long run. Second, the coordination and cooperation within the power and energy value chain can be enhanced and strengthened via sharing of data and knowledge. Integrated business models can be built upon a matured energy ecosystem featuring advanced data acquisition infrastructure and machine intelligence.

At the knowledge level, we put emphasis on the transfer of knowledge, as it helps complete the knowledge flow from locally discovered knowledge to the smart grid knowledge system. Different participants within the smart grid community would benefit from such a knowledge system while actively contributing to it. In Table I, we provide some examples of data, knowledge, and values associated with the participants.

The remainder of this section will focus on explaining some of the critical issues in extending the learning methodologies for smart grid applications. More specifically, the machine learning methods used for power system applications are directly related to the boxes "Data Processing," "Learning Systems," "Big Data Analytic Tools," and "Applications" in Fig. 2. In a broader sense, however, these methods are also highly associated with the closed loops of data and knowledge. It should be noted that learning-based data analytics is not limited to building machine learning models, but also includes using the models to gain insights and new knowledge.

C. Extending the Learning Methodologies for Transparent Power Grids

The past few decades have witnessed the development of machine learning in both theoretical and practical perspectives. Decision trees (DT), artificial neural networks (ANN), support vector machines (SVM) and various other models

TABLE I
EXAMPLES OF DATA, KNOWLEDGE, AND VALUES ASSOCIATED WITH SMART GRID PARTICIPANTS WHEN THE CLOSED LOOPS OF DATA AND KNOWLEDGE ARE FORMED

Participant	Data	Knowledge	Value
Manufacturing companies	Operating data and fault data of power equipment	• Various operating conditions of power equipment can be summarized • Critical pre-fault signals can be identified, and fault-related design flaws can be discovered	• Significant savings in operation and maintenance costs and reductions in losses due to equipment failure or unnecessary scheduled maintenance • Reductions in the loss of information between equipment and managers. • Accelerated product upgrades supported by data
Energy suppliers	• Spatio-temporal data related to renewable energy generation • Real-time power demand data	• Grasp the spatio-temporal dependence of renewable energy generation at multiple scales • Grasp the spatio-temporal distribution of power consumption characteristics and influencing factors	• Improved forecast accuracy and integration capacity of renewable energy • Coordinated supply and demand of energy and improved efficiency of the energy supply chain
Grid operators	Data related to power system operation	• Understand the general patterns of power system operation • Gain the ability to accurately and efficiently analyze and predict the dynamics of power grids at different time scales	• Enhanced safety and reliability of the power grid • Enhanced effectiveness of power system economic dispatch
Electricity retailers	• The electricity consumption data • Data related to demand response participation and results	• Establish accurate electricity consumption models and response models of users • Obtain detailed and accurate user portraits	• More customized services • Improved success rate and effectiveness of demand response • Increased number of customers and revenue
End-users	The electricity consumption data of users	• Gain a quantitative understanding of the electricity consumption characteristics of the household, the neighbourhood, and even larger areas • Have a clear idea of how power consumption and carbon emission of the household can be reduced	• Reduced cost on power consumption • Improved lifestyle with reasonable consumption of power and reduced carbon emission
Data solution providers	First-hand data from various sources	• Be able to have more intuitive and in-depth understanding of various businesses related to power systems • Be more familiar with the characteristics and structures of data in power systems	• Allowing the procedures of data acquisition, storage, and analysis to adapt to the characteristics of power systems • Enhanced competitive power of the companies in the field of power systems • Accelerated development of data science for power system applications
Research institutions	First-hand data from various sources	• Understand the application scenarios and requirements for data science researches in the field of power systems • Understand how domain knowledge can help build data analytics and machine learning models	• Accelerated landing of scientific research for power system applications • Development of interdisciplinary research involving data science, machine learning and power systems

have been intensively studied and used in a variety of industry applications. As a combination of graph theory and probability theory, probablistic graphical models (PGM) have enjoyed much interest in the past three decades for their high representation capability and their learning ability in large networks [13]. Examples of commonly used PGMs include Bayesian networks (BN), hidden Markov models (HMM), Markov random fields (MRF), and conditional random fields (CRF) (readers who are interested in PGMs are referred to [14]). As an important branch of machine learning, Bayesian machine learning has also developed a great deal in the past 20 years [15]. Recently, the implementation of deep learning and RL in electronic games and the game of Go has gained much public attention [16]–[18].

In Table II, we list some examples of power system applications of three machine learning categories, namely, supervised learning, unsupervised learning, and RL. We need to understand that using machine learning methods in power systems is not a new idea. For instance, review papers of applications of ANNs in power systems were published back in 1997 [19]. Thus, instead of thoroughly reviewing existing machine learning methods and their applications, we present some relatively new research results that could potentially inspire researchers in the fields of power systems and point out the reasons how these research results can benefit the development of smart grids in the future.

1) Supervised Learning

In machine learning, we generally find a predictive function based on the data we have. In the supervised setting, we want to find an $f : \mathcal{X} \rightarrow \mathcal{Y}$, where $\mathcal{X}$ is the input space and $\mathcal{Y}$ is the output (label) space. From a probabilistic point of view, many supervised machine learning models can be categorized as either discriminative models (i.e., learning the conditional probability distribution, $P(y|x)$, of the input x and the label y) or generative models (i.e., learning the joint probability distribution, $P(x, y)$) [20]. As labels are readily available for the data, the objective of learning is clear and the result of learning can be easily compared. Thus, supervised learning is by far the most common task in the field of machine learning. Generally speaking, a supervised learning task can be classified as either a classification or a regression task, which are both commonly seen in power systems. Commonly used supervised learning methods include but are not limited to SVMs, ANNs, DTs, random forests (RF), logistic regression, naive Bayes classifiers, k-nearest neighbors (KNN) classifiers and regressors, Gaussian processes, regularized linear regression models, etc [21]. Various signal processing methods including Fourier transform (FT), wavelet transform (WT), S-transform (ST), Hilbert-Huang transform (HHT), Kalman filtering (KF) are often used to extract useful features from raw data before the supervised learning methods are implemented for tasks including fault diagnosis, power quality analysis, and load forecasting [22]–[24].

TABLE II
EXAMPLES OF POWER SYSTEM APPLICATIONS FOR DIFFERENT MACHINE LEARNING CATEGORIES

Category	Applications
Supervised learning	Renewable energy forecasting, power system control, power system stability analysis, load forecasting, demand response, fault diagnosis for transmission lines and distribution systems, non-intrusive load monitoring, power equipment fault diagnosis, electricity market forecasting, electricity theft detection, false data injection detection, power system security assessment, power quality analysis, power system state estimation.
Unsupervised learning	Renewable energy analysis, power system stability analysis, demand response, load profiling, non-intrusive load monitoring, false data injection detection.
Reinforcement learning	Power system control, demand response, electricity market operation, power system economic dispatch.

In recent years, researchers have been showing growing interest in deep neural networks (DNN) for their exceptional performance in certain tasks and unprecedented flexibility, and the learning task fulfilled by which is often referred to as deep learning (note that neural networks are not limited to supervised learning). The implementation of DNNs allows different levels of representations to be learned from data, transforming the process of feature engineering in traditional machine learning pipelines [25]. Applying deep learning to tasks in power systems has just started gaining attention, as the majority of the DNN models are proposed by researchers who focus on tasks including computer vision [26], [27], speech recognition [28], natural language processing [29], [30], etc., for which large amounts of high-quality data can be generated and shared [31]. For applications in power systems, however, accumulating high-quality data in large volumes is no easy task. While we hold a cautiously optimistic attitude towards implementing DNNs in power systems, we do not put too much emphasis on the depth of the neural network models in this paper. Examples of neural network models for applications in transparent power grids are listed in Table III. Two building blocks of neural networks, namely, convolutional neural networks (CNN), and recurrent neural networks (RNN) are mentioned in the table. Generally speaking, CNNs are often used to extract spatial features, and both CNNs and RNNs can be used to extract temporal features or model temporal dependence.

An obstacle for implementing standard neural network models is that the signals collected by sensors (e.g., voltage and current amplitudes) in power systems naturally reside on nodes of a network. In [39], the basic concepts of signal processing on graphs are introduced, and operators including filtering, convolution, and translation in graph setting are reviewed in detail. This provides researchers with a new perspective for analyzing data generated in power systems. More specifically, some recent works focus on the implementation of CNN models on graphs [40], [41], and both spatial and spectral constructions of graph convolutional layers are established. These layers can then be incorporated into more complex neural network models. In addition, the deployment of WSNs in the power networks will greatly facilitate the implementation of such methods, as the ubiquitous sensors in the WSNs would enhance the observability of the signals on all nodes of the power networks. In Fig. 3, an illustrative DNN model with spatial graph filters (see [40] for the detailed construction) and

TABLE III
EXAMPLES OF NEURAL NETWORK MODELS FOR APPLICATIONS IN TRANSPARENT POWER GRIDS

Nature of the Learning Task	Model Structure	Examples of Application
Temporal	Temporal feature extraction (CNN)	Fault detection and classification [32] Residential demand response [33]
	Temporal dependence modeling (RNN)	Generating power forecasting for photovoltaic system [34] Residential load forecasting [35]
	Temporal feature extraction (CNN) + temporal dependence modeling (RNN)	Non-intrusive load monitoring [36] Short-term wind power forecasting [37] Short-term load forecasting [38]
Spatial	Spatial feature extraction (CNN)	Power system transient stability analysis Anomaly detection in microgrids
Spatio-temporal	Spatial feature extraction + temporal feature extraction (CNN)	Power system fault location Power system steady state stability analysis
	Spatial feature extraction (CNN) + temporal dependence modeling (RNN)	Forecasting of renewable energy using numerical weather prediction Charge scheduling of electric vehicles

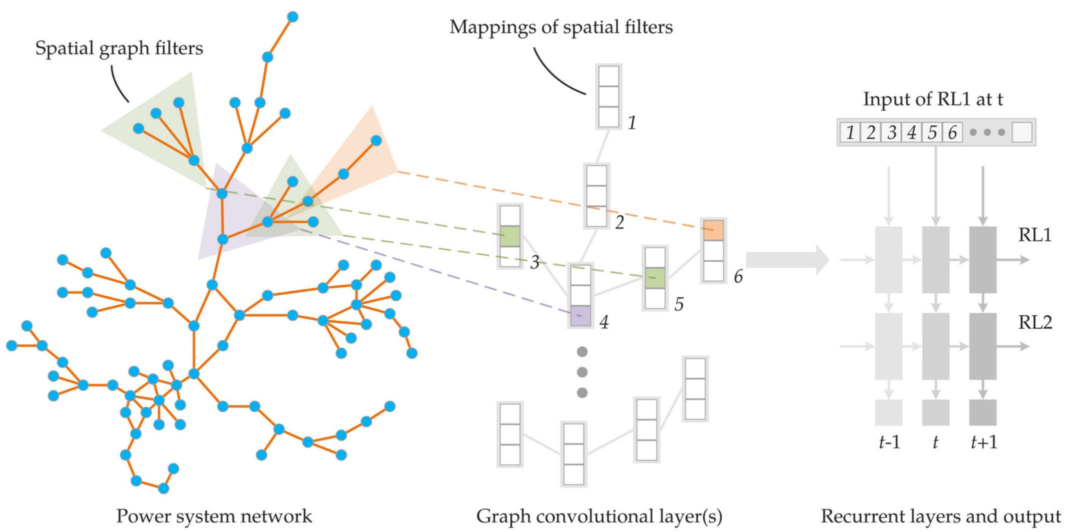


Fig. 3. An illustration of a DNN model with spatial graph convolutional layers and recurrent layers. In this illustrative model, there are three spatial graph filters marked with different colors. The mappings of these filters of the same time step are used to form the input vector of the recurrent layers. A total of two recurrent layers are placed before the output of the model.

recurrent layers is demonstrated. A more practical construction of the graph filters is the spectral construction [40], [41], which can be implemented by replacing the spatial convolutional layer in Fig. 3 correspondingly.

Several paths for developing supervised learning methods for power system applications include:

1) Finding ways to obtain useful models using imbalanced datasets. Imbalanced data is unavoidable for many power system applications. For instance, the chance of observing electrical equipment failures may be very low due to scheduled maintenance, and certain types of faults may be less likely to occur in power systems. Commonly used methods for imbalanced learning include sampling methods, ensemble methods (e.g., bagging and boosting), cost-sensitive methods, kernel-based methods, and active learning methods [42].

2) Finding ways to learn effectively from small datasets. A practical way to deal with machine learning tasks with small datasets is to look for inconspicuous but useful features and hand craft effective machine learning models, which requires domain knowledge and rich experience. Unsupervised learning and transfer learning, two concepts that will be introduced later in this paper, can also be very helpful when the dataset used for a learning task is relatively small.

3) Finding ways to deal with spatio-temporal data, especially signals residing on graphs (e.g., transmission networks, distribution networks, and transportation networks that are coupled with power systems).

4) Dealing with the uncertainties (e.g., the uncertainties of weather dynamics or human behavior) in power systems.

5) Designing neural network structures suitable for various data sources and applications in power systems. Apart from combining different building blocks into a single model, another path worth exploring is adding reasoning, attention, and memory to the models so that both long-term dependence and short-term context can be fully considered [29], [43]–[45].

6) Increasing the explainability and interpretability of machine learning models used for power system applications.

2) Unsupervised Learning

Unsupervised learning generally refers to the learning tasks that learn from unlabeled data. Examples of unsupervised learning tasks include clustering, anomaly detection, dimensionality reduction (e.g., principal component analysis (PCA), and independent component analysis (ICA)), association rules analysis, graph structure discovery, etc. (for a detailed introduction to unsupervised learning, the readers are referred to [21], [46]). The most common unsupervised learning task is to cluster unlabeled data samples. Algorithms or models used for clustering include k-means, hierarchical clustering, spectral clustering, Gaussian mixture models (GMM), Dirichlet process mixture models (DPMM), self-organizing maps (SOM), density-based spatial clustering of applications with noise (DBSCAN), etc. A review of clustering methods for load pattern grouping can be found in [47]. Recently, some active research topics of unsupervised learning include learning representations from data, and generative models in the unsupervised setting:

1) *Learning representations from data:* A comprehensive review of representation learning is provided in [48]. Generally speaking, learning representations reveals the inherent characteristics of the data samples within the dataset being studied, and complex tasks can be completed on the basis of these representations. From a single-layer learning module perspective, two representation learning paradigms can be identified, one focused on PGMs (e.g., restricted Boltzmann machines (RBM)), and the other one focused on learning direct encodings (e.g., autoencoders) [48]. Both RBMs and autoencoders can be used to build DNNs in a layerwise manner (though recent research has revealed that pre-training is not necessary). Nevertheless, the end-to-end nature of DNN models is also partially explained by the belief that expressive representations can be learned by the layers within the networks. Some recent works in the field of power systems highlighted the importance of representation learning. In [49], sparse autoencoders are used to learn representations from overvoltage datasets. Unsupervised feature learning is also used in [32] to extract features from voltage and current waveforms related to faults in transmission lines. The features are then used in a CNN which detects and classifies the faults. Non-negative K-SVD is used to learn representations of daily load profiles in [50]. Some other examples can be found in Table III [33], [36], [37].

2) *Generative models in the unsupervised setting:* Two types of generative models have emerged recently and gained much attention, namely, variational autoencoders (VAE) [51], [52], and generative adversarial networks (GAN) [53]. Specifically, VAEs are able to generate a data point x using $p_\theta(x|z) = f(z)$, where z is the latent variable and $p_\theta(x|z)$ is the likelihood with generative model parameters θ. In case the true posterior density $p_\theta(z|x)$ is intractable, it is approximated by $q_\phi(x|z)$. From the point of view of coding theory, $q_\phi(x|z)$ can be referred to as an encoder, and $p_\theta(x|z)$ can be referred to as a decoder [51]. For a VAE, both the encoder and the decoder are learned by training two neural networks with backpropagation (the prior of the latent variables is set to $N(0, I)$). For a GAN, two models, namely, a generative model, G, and a discriminative model, D, are trained at the same time via a two-player minimax game [53]. The generative model is able to generate data samples that are likely to be sampled from the distribution of the training data using random noise in the latent space, while the discriminative model is expected to tell whether a data is from the training data or G. When both D and G are neural networks, they can be trained by backpropagation, after which G is able to generate realistic data samples. The generative models can be extended to the semi-supervised setting, where limited labeled data is available (which is true for a lot of applications in power systems) [54]. The above-mentioned generative models are quite promising to be used for power system applications, especially for renewable energy generation and demand-side applications, where we not only want to capture information from existing data but also want to generate data samples that are unobserved but likely to be seen in the real world. In [55], the authors use GAN to generate wind and photovoltaic power profiles with great diversity. The generated realistic profiles can help capture the variations and uncertainties in renewable energy generation. The ability to generate data samples can also be used in predicting various quantities (e.g., short-term energy consumption prediction).

In summary, we believe that unsupervised learning for power system applications will be manifested in the following ways:

1) Learning critical characteristics and patterns of the dynamics of power systems;

2) learning reliable and useful data representations for spatio-temporal data in power systems;

3) learning the internal models, structures, and regularities of complex systems (e.g., power grids, electricity markets);

4) generating realistic and useful data using generative models.

These implementations of unsupervised learning are of great significance for building transparent power grids, as we can learn from the huge amounts of unlabeled data and gain a better understanding of the intrinsic characteristics of the various components of power systems.

3) Reinforcement Learning

The learning in power systems goes beyond the settings of supervised learning and unsupervised learning. System operators interact with the system by monitoring the status of the system and taking a series of actions accordingly. The feedback of the actions, however, may not relate directly to the objective of taking these actions in the first place. Thus, in order to develop an intelligent system operator, the learning agent needs to explore an optimized policy (the action-selection strategy) while trying different actions and getting feedback from the environment, so that the actions to be made would lead to maximized expected total reward.

Concretely, RL is typically formulated as a Markov decision process (MDP) as follows [56]. Let $s_t \in \mathcal{S}$ be the state that can be observed by the agent at time t in the environment E, and let r_t denote the reward at time t. $P^a_{s\rightarrow s'} = P(s'|s, a)$ is the transition function associating s and s', and $R^a_{s\rightarrow s'}$ is the reward function, where $a \in \mathcal{A}$ is the action chosen by the agent. The goal of the agent is to collect as much reward as possible by finding an optimal policy $\pi(s, a) = P(a_t =$

$a|s_t = s)$ (which can also be deterministic), that is,

$$\pi^* = \underset{\pi}{\operatorname{argmax}} \sum_{s\in\mathcal{S}} V(s) \tag{1}$$

where π^* is the optimal policy, and $V(s)$ is the value function which evaluates the expected return for π. More specifically, for a continuing task, $V(s) = \mathbb{E}_\pi[\sum_{t=0}^{+\infty} \gamma^t r_{t+1}|s_0 = s]$, where $\gamma \in [0, 1)$ is the discount factor. In addition, the Q function, which evaluates the expected return when action a is chosen for s, is defined as $Q(s, a) = \mathbb{E}_\pi[\sum_{t=0}^{+\infty} \gamma^t r_{t+1}|s_0 = s, a_0 = a]$. When the model behind the task is known (i.e., $R^a_{s\to s'}$ and $P^a_{s\to s'} = P(s'|s, a)$ are known), the optimal policy can be obtained by policy iteration or value iteration [56]. In more practical settings where the model is unknown, model-free learning methods such as Q-learning [57] are often used. In Q-learning, the rule of updating the Q function is:

$$Q(s, a) := Q(s, a) + \alpha(r + \gamma \max_{a'} Q(s', a') - Q(s, a)) \tag{2}$$

where r is the return for choosing action a at s, and s' is the subsequent state [58].

Researchers have been using RL in a variety of applications related to power and energy systems (see [59] for an extensive literature review). Three topics, namely, residential demand response, power system control, and electricity market are good examples of applying RL in power systems:

1) *Residential demand response:* In [60], an optimal policy is learned based on energy prices and consumer reservations using Q-learning, such that the total discounted cost can be minimized. The cost function in this paper consists of the financial cost and a pre-defined dis-utility cost which measures the dissatisfaction of consumers when their reservations are delayed or only a proportion or required energy is allocated. In order to reflect different consumers' characteristics of dissatisfaction, authors of [61] proposed a method that learns a consumer's dissatisfaction based on his evaluations on response jobs (either completed or canceled). In addition, the jobs initiated by the energy management system are also considered. In [62], instead of modeling the energy price as a random process [60], [61], the authors proposed a model in which the service provider needs to dynamically decide the retail pricing function. Multi-agent Q-learning is used, as both the actions of service providers (dynamic pricing) and consumers (consumption scheduling) are considered. The fitted Q-iteration proposed in [63] is extended in [64] in order to use the forecast of exogenous data. In [33], a CNN is used to estimate the Q function in the supervised learning step of the fitted Q-iteration. State-time features can be automatically extracted by the CNN when a sequence of past observations is used as the input. As a matter of fact, combining DNNs and RL has been a recent promising topic in the field of machine learning, which greatly extends the way in which observations from the environment are handled. The multi-agent smart home automation system proposed in [65] provides a systematic framework for implementing RL in households based on WSNs.

2) *Power system control:* The nature of RL makes it highly suitable for various control tasks in the power system. The framework for power system stability control based on RL is introduced in [66], in which two learning modes, namely, the online mode (the agent interacts with a real power system) and the offline mode (the agent interacts with a simulation model), are considered. Q-learning is applied to reactive power control in [67]. In [68], the authors compared fitted Q-iteration with model predictive control in a benchmark power system oscillation damping problem and showed that the fitted Q-iteration is able to give robust control policies with satisfactory accuracy. The authors also pointed out that "offline global (fitted Q-iteration)" and "online local (model predictive control)" approaches can be combined to achieve better results. The problem of load frequency control is solved using multi-agent RL in [69], and generic algorithm is used for parameter-tuning. Wide-area measurement data is used in [70] to train the agent for power system stability control. Decentralized controllers are fed with data pre-processed by the central unit, and each decentralized controller calculates its local control signal in an autonomous manner and applies it to the controlled devices nearby.

3) *Electricity market:* RL is used to seek Nash equilibria of energy trading games [71], [72]. In [73], the electric vehicle (EV) fleet charging schedule is determined using RL in a day-ahead electricity market. Cost-effective day-ahead plans for aggregators can be learned using readily available EV charging parameters.

A natural extension for building RL models is to incorporate the building blocks of deep learning, which is particularly helpful for large state spaces or action spaces (interested readers are referred to [74], [75]). In addition to the perspectives for RL in [59], we further put forward several key issues for the implementation of RL in power systems:

a) The cooperation of multiple agents designed for different control tasks is required in order to ensure the safety and reliability of the power systems. Thus, a coordinator is needed to make sure that the actions taken by different agents are compatible.
b) Physical models and constraints are of great significance for RL in power systems, especially for the simulation of the systems. It is thus of great interest to design a framework for power system simulation which is able to produce accurate predictions given certain actions taken by the agents in a reasonably short time.
c) One way to integrate domain knowledge and safety concerns is to add human feedback to the RL framework [76]. This would also make sure that the experience of human experts can be utilized in the evaluation procedure of the RL model.

In consideration of the above-mentioned concerns, we propose a modified scheme for implementing RL in the transparent power grid. In Fig. 4, we illustrate the proposed scheme based on the actor-critic framework [77]. The actors are designed for different tasks and have individual critics. Note that the simulator is not limited to traditional power system simulation systems (e.g., it may include a hierarchy of statistical models that are designed with considerations of empirical prior knowledge and physical constraints). Human experts can send their feedback with respect to the performance of the RL system to the reward predictor so that human preferences and

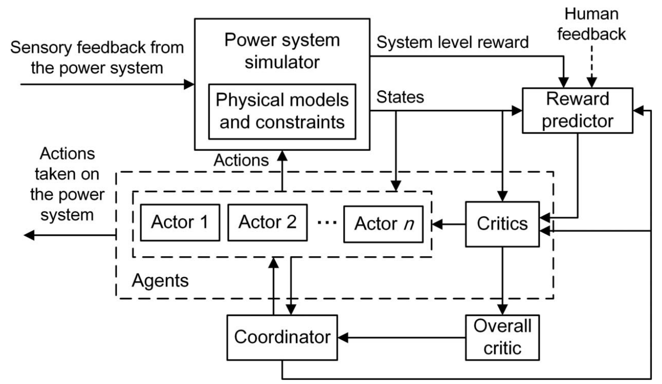


Fig. 4. A diagram of the proposed reinforcement learning scheme for power systems. Human feedback is used for the reward predictor to fit its reward functions.

concerns can be reflected in the rewards sent to the critics. The safety of the power systems can be ensured via the physical constraints, the human feedback, and the online RL strategy based on predictions provided by the simulator.

4) Some Principles and Best Practices of Learning

Implementing a suitable learning model in an appropriate way is not an easy task. Prior to actually carrying out a practical project, several principles and best practices of learning from data merit our attention:

1) *The complexity of the model and the amount of data:* Choosing a suitable model for a particular learning task is a crucial step to take. Two factors need to be considered: the complexity of the model and the amount of data available. Generally speaking, the complexity of a model decides how well the model can fit the data, and higher complexity means increased fitting ability. According to the Vapnik-Chervonenkis (VC) theory, however, high model complexity leads to raised expected errors which measure the generalizability of the model [78]. This can be further explained by the bias-variance tradeoff by focusing on only the cases of (high bias, low variance) and (low bias, high variance): a simple model may have stable performance (low variance) but may also fail to fit the training dataset well, while a complex model may easily fit any given training dataset (and is thus able to find the best fit for the true data distribution when the whole distribution is known) but is prone to give unstable results for multiple training datasets (high variance). The latter case is often referred to as over-fitting (which is often dealt with by cross-validation, regularization, early stopping, etc., when the dataset and the type of model are fixed). Thus, it is favorable to choose the simplest model that fits the data [79], such that both bias and variance can be kept relatively low at the same time. Meanwhile, it is desirable to obtain more data when a better performance of the model is required. The recent success of DNN models relies heavily on the amount of data available. When the amount of data available is limited (which is the case for many power system applications), an alternative is to implement dataset augmentation to create new fake data [31]. In [80], the authors leverage the information contained within unlabeled and partially labeled data to improve the performance of an event-detection system in power distribution networks. This semi-supervised learning setting is applicable to a large number of data-driven tasks in power systems.

2) *The proper usage of data:* It is necessary to separate the dataset available into a training set and a test set so that the performance of the learned model can be evaluated on some metrics. In order to improve the learning process (e.g., tune the hyper-parameters or change the learning algorithm), a validation set can also be used. In [81], the author suggests that, instead of randomly dividing the dataset into three sets, a better approach is to ensure that the validation and test sets reflect data that is more likely to be seen in the future. In addition, the distribution of the validation set and the test set need to be the same, as only the validation set is used during the development stage. As a matter of fact, it is very important to not look at the test set during the whole procedure of developing the model. A more strict requirement is to choose the learning model before looking at the data closely [79]. For some applications, it is infeasible to collect a lot of data (e.g., failure of rotating machines). It is then very important to keep in mind that designing the learning model after looking at all the data samples or using the whole dataset to tune the hyper-parameters should be avoided. When a very large dataset is available, the sizes of the validation and test sets can be properly controlled as long as they are large enough to evaluate the performance of the model [81].

3) *Ensemble learning for improved overall performance:* It is practical to combine a number of models to boost the overall performance in a specific learning or data mining task. Existing ensemble methods include bootstrap, bagging (e.g., random forest), boosting (e.g., AdaBoost and gradient boosting machines (GBM)) [82], etc. The ensemble approach is well suited for the development of practical machine learning projects in power systems and can be easily extended to different application scenarios in order to obtain satisfactory performance.

III. Challenges for the Closed Loops of Data and Knowledge

Building the closed loops of data and knowledge is a difficult task to complete. The first major challenge lies in the protection of security and privacy against internal and external attacks. The transfer of knowledge is also a challenging task, as existing data mining and machine learning methods and models are not specifically designed for knowledge transferring. In this section, we survey existing solutions to the challenges faced by the closed loops of data and knowledge towards transparent power grids and point out some paths for future works.

A. Privacy-preserving Data Analysis for Smart Grid Applications

It should be noted that in some cases we need to trade off between privacy and performance. In [1], several critical aspects related to energy big data security and privacy is discussed and some promising studies are highlighted [9], [83]–[85]. Confidentiality, integrity, authenticity, non-repudiation, utility, scalability and cost are used to evaluate the functionality of various cryptosystems [1]. With growing concern over possible cyber-attacks against WSNs and the AMI, two important issues merit attention. First, side information or prior knowledge may be used by attackers to infer the metering information of users. Further, the loss of utility and usability of data under privacy-preserving schemes needs to be quantified and minimized when revealing summary statistics, synthetic datasets, machine learning models, etc. The concept of differential privacy, which is proposed by Dwork *et al.* [86], [87], is able to address the two issues and thus is worth in-depth discussion. Concretely, an algorithm $\mathcal{A}$ is $\epsilon-$differential private, if for all pairs of datasets D_1 and D_2 at most differing in the data of one single user, and for all $S \in range(\mathcal{A})$,

$$Pr(\mathcal{A}(D_1) \in S) \leq e^{\epsilon} \cdot Pr(\mathcal{A}(D_2) \in S). \quad (3)$$

The quantitative trade-offs between privacy and statistical efficiency can be further obtained using minimax rates, as introduced in [88], [89], where local differential privacy is preserved if data is kept private even from the statistician or learner. More formally, $X_1, \cdots, X_n \in \mathcal{X}$ are samples drawn from some distribution $P \in \mathcal{P}$, and the task is to estimate the parameter $\theta := \theta(P)$ of the distribution with access only to $Z_1, \cdots, Z_n \in \mathcal{Z}$, which are obscured views of the original samples. These obscured views are obtained via a conditional distribution $Q \in \mathcal{Q}_\alpha$, where α is the privacy parameter which controls the level of privacy. The parameter $\theta \in \Theta$ is estimated by $\widehat{\theta}$ and the quality of the estimation is evaluated by $\mathbb{E}_{P,Q}[\ell(\widehat{\theta}(Z_1^n), \theta(P))]$, where ℓ measures the error of estimation by $\widehat{\theta}$. The $\alpha-$minimax rate for the family $\theta(\mathcal{P})$ is defined as

$$\mathfrak{M}_n(\theta(\mathcal{P}), \ell, \alpha) := \inf_{\widehat{\theta}, Q \in \mathcal{Q}_\alpha} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q}\left[\ell(\widehat{\theta}(Z_1^n), \theta(P))\right]. \quad (4)$$

Sharp bounds on convergence rates for many standard statistical inference procedures under local differential privacy can then be developed and explored [89].

In [90], possible applications of differential privacy in signal processing and machine learning is discussed. In addition to a brief introduction to signal processing and machine learning with privacy, three topics of interest are raised. First, in order to integrate differential privacy into signal processing applications, the design of signal acquisition needs modification. Second, more work needs to be done for core signal processing tasks including prediction, forecasting, signal transforms, especially with more complex signals. Third, networked information systems with parties who are willing to collaborate without divulging local data provide a lot of potential applications. All these issues are closely related to the closed loops of data and knowledge in transparent power grids. More recently, the combination of differential privacy with deep learning is gaining attention [91]. More specifically, a differentially private training approach of neural networks based on differentially private stochastic gradient descent (SGD) is proposed.

Several studies have used differential privacy in the analysis of smart metering data [92]–[96]. In [92], the authors discuss the privacy-preserving data aggregation scheme involving a number of smart meters and a utility company. The proposed approach in the paper, which contains a differentially private algorithm, only allows the utility to have access to aggregated consumption data containing Laplacian noise. Thus, the privacy of single meters can be strongly protected against the utility company. The authors of [93] propose a differentially private battery-based load hiding (BLH) method against non-intrusive load monitoring (NILM). Certain differential privacy bound can be achieved without violating the constraints of the batteries, so that the behavior patterns within a household cannot be revealed. The BLH problem is extended in [94], where the authors discuss how differential privacy can be achieved with battery recharging. In [95], the authors propose a privacy-preserving metering aggregation scheme against human-factor-aware differential aggregation (HDA) attack, which tries to infer the activities of a household by comparing aggregate consumption in different time slots. A fair data sharing mechanism based on differential privacy is proposed in [96], where the consumers are compensated for their participation in the data market (increased participation indicates increased information leakage). Differential privacy for smart meters is discussed in [97], where comparisons with other methods are provided. Other promising privacy-preserving mechanisms for data analytics in smart grids include homomorphic encryption [98], multi-party computation [99], and trusted computing [100].

Some open issues related to data privacy in a transparent power grid include:

1) *The privacy-preserving mechanism of data sharing:* while it is well acknowledged that an increased amount of data leads to improved performance of machine learning models, the direct sharing of data may be unrealistic due to practical issues and restrictions. This calls for mechanisms that enables utilization of data from multiple parties without data exchange between any two parties. For instance, in [101], the author introduces a distributed learning approach for wind power forecasting that uses wind power generation data from multiple wind farms without direct sharing of data.

2) *The proprietary and exploitation rights of privacy data:* in an advanced electricity market, various parties including distribution companies, retailers, market operators, independent system operators, regulators, and consumers are highly related and produce large amounts of data among them. It is thus of great urgency and significance to clarify each party's rights and obligations with respect to various types of data from different sources within the market. Each party should be entitled to know how the data it owns is distributed and used by other parties, especially the parties that are not directly related to it but may make profit from the data.

3) *The trade-off between customization and breach of privacy:* data-driven customization simplifies customers' lives to a great extent in many ways. While this can possibly be true for residential electricity consumers who take part in demand response programs, they might put higher priority on privacy. As different customers have varied tolerances for privacy violations, service providers need to come up with a variety of scenarios that cater to the preferences of different customer groups. High transparency of data collection and usage should also be guaranteed.

B. Transfer of Knowledge and Experience Across the Smart Grid Ecosystem

One of the core concepts in the closed loop of knowledge is that knowledge can be transferred from some domains or tasks to others domains or tasks. The transfer of knowledge is of special significance for industry applications, as collecting or labeling data may be costly or even infeasible in many cases.

In the field of machine learning, the transfer of knowledge is often referred to as transfer learning. A comprehensive survey on transfer learning is provided in [102]. Generally speaking, traditional data mining and machine learning algorithms are implemented under the condition that the data used in training and testing (whether labeled or not) follows the same distribution. Transfer learning, however, allows the distribution of data as well as domains and tasks used in training and testing to be different. More specifically, let a specific domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$, where $X = \{x_1, \cdots, x_n\} \in \mathcal{X}$, and $P(X)$ is the marginal probability distribution. Also, let a specific task $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$, where $\mathcal{Y}$ is the label space, and $f : \mathcal{X} \to \mathcal{Y}$ is used to predict the corresponding label of a given $x_i \in \mathcal{X}$ whose true label is $y_i \in \mathcal{Y}$. Given domains $\mathcal{D}_S$ and $\mathcal{D}_T$, learning tasks $\mathcal{T}_S$ and $\mathcal{T}_T$, transfer learning is aimed to help improve the learning of $f_T(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_T$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$ [102]. Concretely, transfer learning can be further categorized under three settings, namely, inductive transfer learning, transductive transfer learning, and unsupervised transfer learning depending on the availability of labels [102]:

1) *Inductive transfer learning:* In the inductive transfer learning setting, the labels of the target domain are available. When labeled data in the source domain is also available, the source domain can construct useful feature representations in a supervised manner. For instance, the fault location and fault classification tasks for power distribution systems can share learned features of either one of the tasks. The forecasting of renewable energy generation can also benefit from sharing the forecasting model, as multiple farms are usually spatially correlated and have similar characteristics. In [103], several wind farms share the hidden layers of a DNN model. The results show that the transfer of data among wind farms is helpful, especially for farms with small datasets.

When no labeled data is available in the source domain, the setting of inductive transfer learning is similar to that of self-taught learning, as introduced in [104]. Unsupervised feature construction methods such as sparse coding can be used to extract features of load data and thus facilitate the prediction or demand response tasks [50].

2) *Transductive transfer learning:* In the transductive transfer learning setting, the labels of the source domain are available, while the labels of the target domain are unavailable. For instance, some knowledge can be transferred between inland and off-shore wind farms for certain tasks, in which case the data may come from the same feature space but have different distributions. The knowledge of fault detection for various types of equipment within the power system can also be shared, especially for faults that rarely occur or often cause catastrophic consequences.

3) *Unsupervised transfer learning:* Unsupervised transfer learning can be implemented when labels of both the source domain and the target domain are unavailable. For instance, the clustering task in the target domain may yield better results with the help of unlabeled data in the source domain.

More specifically, approaches to transfer learning include transfer of instances (parts of datasets), feature representations, parameters, or relational-knowledge (some relationship within the data) [102]. Recent studies on transfer learning are closely related to DNNs due to the fact that the lower layers are prone to learn low-level feature representations, which are likely to be invariant for different but related tasks. The transfer learning can be completed by copying the parameters of the lower layers and fine-tuning the parameters of the higher layers using labeled data in the source domain [105]. Due to the fact that the majority of datasets used for learning tasks are small-sized, it is expected that applying transfer learning to power system applications will effectively improve the performance of the models.

C. Combining Machine Learning and Causal Inference for Power System Applications

The most commonly used machine learning algorithms today are good at modeling the associations between input and output features (e.g., $P(y|x)$ for supervised learning), but a limitation of these algorithms is that they are generally not suitable for causal inference. Recent criticisms of deep learning models also emphasize that they are not good at dealing with causal relationships [106] despite their strong ability to learn representations and mappings. In [107], the author introduces the impediments to machine learning towards causal inference and points out the other two levels in the causal hierarchy above the level of association (i.e., $P(y|x)$):

1) *The level of intervention:* Expressions of this level take the form of $P(y|do(x))$, i.e., the probability of event $Y = y$ when intervention X is set to x, which can be estimated by random trials (experimentally) or Bayesian networks (analytically).

2) *The level of counterfactuals:* Expressions of this level take the form of $P(y_x|x', y')$, i.e., the probability that $Y = y$ would occur had X been x, given we only observe that $Y = y'$ and $X = x'$.

The author points out that model-free statistical learning approaches are insufficient to infer causal relationships. Thus, he proposes a framework of structural causal models (SCM), which contains graphical models, structural equations, and counterfactual and interventional logic. This framework is able to handle the above-mentioned three-level causal hierarchy.

It should be pointed out that causal inference is of great significance to the establishment of the closed loop of knowledge in transparent power grids. Causal effects are very important for a variety of power system applications, as they can greatly facilitate the decision-making process. Take demand response as an example: different incentives may result in varied responses from residential users, and the combination of all users' response results is the overall result of the demand response program. However, it is insufficient to obtain the expected response of a certain user under different demand response plans using only historical data. Let us consider the effect of incentive X on a residential user's electricity consumption Y:

1) At the level of association, we can estimate the expected value of Y at $X = x_0$ from historical data. The estimation, however, is based only on observations from the past and can give misleading results.

2) At the level of intervention, we want to know the expected value of Y if we set X to x_0. Since the conditions at the time when we set x to x_0 are not the same as the conditions when x evolved naturally to x_0 in the past, the user's reaction may be very different.

3) At the level of counterfactuals, we already know that $Y = y_0$ when $X = x_0$ at a particular time $T = t_0$, and we need to calculate the expected value of Y had we set X to x_1 instead of x_0 at $T = t_0$. At this level, we cannot get the answer experimentally because the conditions at $T = t_0$ cannot be reproduced. Achieving this task can help us assess the effectiveness of the demand response programs. The design of the programs can also be more straight-forward and well-directed after analyzing the causal effects at this level. The same discussion may also apply to electricity market policymaking, power system stability analysis, power equipment fault diagnosis, and other similar power system applications.

We can also extend the discussion at the level of intervention to the transportability of causal effects. In [108] and [109], the authors introduce how causal information learned from experiments can be transferred to an environment where only passive observations can be made. For example, suppose we have learned $P(y|do(x), z)$ for a price-based demand response program in city A, where X is the electricity price, Y is the electricity consumption of a certain user, z is the value of an observable variable Z, and we want to find a way to estimate $P^*(y|do(x))$ in city B. Specifically, Z can be a variable that has causal effects on X and Y (e.g., household income, number of family members, etc.), a proxy for such variables, or an X-dependent variable on the causal pathway between X and Y. By analyzing the different conditions of Z, we can determine what experiments and observations are needed and how to combine them into the estimation of $P^*(y|do(x))$ [110]. Note that this is also related to the aforementioned task of transfer learning.

Examples of linking machine learning with causal inference also include the implementation of causal trees [111] or causal forests [112], and representation learning by neural networks for counterfactual inference [113]. It is expected that the in-depth understanding of causality inference can make machine learning systems more understandable, explainable, and accessible for power system applications.

D. Scalable and Flexible Data Analytics for Applications in the Power and Energy Industry

A major challenge faced by data analytics in the power and energy industry is that a large proportion of collected data is heterogeneous [12]. That is, various types or formats of data may coexist in the database and may need to be handled simultaneously. Although we primarily focus on the field of machine learning in this paper, we need to note that addressing the storage and processing of heterogeneous datasets is also crucial for other data mining tasks as well as business intelligence applications in the era of big data.

At the enterprise level, data curation becomes a critical issue, as the management of data would be constantly challenged by data sources and applications that are both changing rapidly. The data management system between data sources and applications needs to be able to scale horizontally by defining unified interfaces, such that new data sources and applications can be integrated without repetitive work. For real-time applications (e.g., demand response or electricity market operations), a high-throughput, low-latency data system becomes indispensable.

E. Efficient and Reliable Computation for Learning and Data Mining

It is widely acknowledged that more data (with high quality) means better performance for machine learning and data mining in complex tasks. While the learning and mining processes may not run online, the increasing volumes of data still pose a huge burden on the the computation resources.

Distributed computing is critical to fast, efficient and reliable computation. Higher-level tasks including Spark's scalable machine learning can run on top of work nodes in the distributed file system (e.g., Hadoop distributed file system, HDFS) with the help of advanced resource managers (e.g., yet another resource negotiator (YARN)) [12]. With large volumes of data being collected by dispersed sensors and different levels of applications being deployed in the smart grid hierarchy, cloud-based big data solutions are gaining attention. In [9], the authors proposed a cloud-computing-based information management framework for smart grids. The relationships among different roles in the cloud computing model for smart grids is discussed in [114]. Three types of networks, namely, private networks (established by utility companies), local area networks (established by customers), and public internet, are included in the proposed cloud-computing framework in [115]. An identity-based security scheme is also proposed to deal with the security threats posed by public cloud services and data transmission over the internet.

For a variety of algorithms and applications, the key to fast implementation is to adopt parallel computations. For instance, recently, graphical processing units (GPU) are widely used in both academia and industry to train DNNs with huge amounts of parameters, thanks to the fact that neural networks are inherently parallel. In [116], the amounts of time used for state estimation tasks completed by CPUs and GPUs are compared. Results show that the task runs much faster on GPUs than CPUs when the number of buses in the system, and the ratio between them enlarges when the number of buses in the power system increases.

IV. Integrated Solutions Featuring the Industrial Internet and the Digital Grid

The concepts of the Internet of Things (IoT) and the Industrial Internet have received much attention from a variety of industries, as the interconnection between objects within the physical world requires a high-quality and reliable data collection and transmission mechanism. Recently, companies including General Electric (GE) and Siemens are proposing their own solutions concerning the digital transformation of power and energy systems [117], [118]. Although the solutions provided by different companies differ in their emphases and features, the value of data analytics is given unprecedented significance in all of the solutions.

General Electric (GE) introduced the concept of the Industrial Internet as the new wave of innovation in 2012 [119]. The need of analyzing and understanding industrial systems at multiple levels as well as the development of internet-based innovations has inevitably given rise to the integration of intelligent machines, advanced analytics and people at work, laying the foundation of the Industrial Internet. With the help of low-cost sensors, data of everything from generators to transformers on power poles can be collected, thus keeping operators well-informed of possible outages and saving the time and effort of field representatives. The digital transformation of the power industry was further discussed in [117] with special emphasis laid on the impact of the Industrial Internet. The new energy value chain will be established based on digital technologies and the interconnection of numerous critical assets. More specifically, it is estimated by GE that over seven billion devices will be installed by 2020 across the energy value chain with Exabytes of sensor data being collected on a yearly basis. GE is also working together with AT&T to merge GE's smart grid solutions with AT&T's IoT infrastructure and the two companies are jointly providing solutions to customers [120]. In addition, GE and Microsoft are reportedly preparing to bring Predix, the Industrial Internet platform of GE, to the Windows Azure platform in order to equip Predix's customers with advanced artificial intelligence, data visualization, and other related techniques [121].

In 2014, Siemens and Accenture formed a joint venture company, OMNETRIC Group, whose core business is providing clients with smart grid solutions and services. The combination of strengths of both Siemens and Accenture arguably enables the OMNETRIC Group to demonstrate their expertise in energy technologies as well as systems integration and operations strategy. Concretely, the company is focused on achieving value from information and operational technology integration [122] and data discovery which makes full use of almost all data available [118].

The integrated solutions provided by the above-mentioned companies reflect the ever-growing importance of data collection by sensors and advanced data analysis techniques. In order to push forward the development of power systems featuring IoT and II, joint efforts of academia and industry are indispensable.

V. Conclusion

In this paper, the concepts of learning-based data analytics moving towards transparent power grids are presented. We first introduce the features of transparent power grids, followed by the closed loops of data and knowledge, which is also of great significance in the upcoming era of the Industrial Internet and digital grids. The framework of data analytics and learning is extended in both vertical and horizontal directions. In particular, we focus on the application of advanced machine learning techniques in power systems. It is hopeful that the data analytic framework presented in this paper will make power grids more transparent, efficient, intelligent, and user-friendly in the future.

References

[1] J. K. Hu and A. V. Vasilakos, "Energy big data analytics and security: challenges and opportunities," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423–2436, 2016.

[2] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 10, pp. 3557–3564, Oct. 2010.

[3] S. Chowdhury and P. Crossley, *Microgrids and Active Distribution Networks*. Scotland: The Institution of Engineering and Technology, 2009.

[4] P. Mancarella, "MES (multi-energy systems): An overview of concepts and evaluation models," *Energy*, vol. 65, pp. 1–17, 2014.

[5] European Council. (2014, Oct.) European Council (23 and 24 October 2014)-conclusions on 2030 climate and energy policy framework. [Online]. Available: http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/145356.pdf

[6] The White House. (2014, Nov.) U.S.-China joint announcement on climate change. [Online]. Available: https://obamawhitehouse.archives.gov/the-press-office/2014/11/11/us-china-joint-announcement-climate-change

[7] Y. Simmhan, S. Aman, A. Kumbhare, R. Y. Liu, S. Stevens, Q. Z. Zhou, and V. Prasanna, "Cloud-based software platform for big data analytics in smart grids," *Computing in Science & Engineering*, vol. 15, no. 4, pp. 38–47, Jul./Aug. 2013.

[8] F. J. Luo, Z. Y. Dong, J. H. Zhao, X. Zhang, W. C. Kong, and Y. Y. Chen, "Enabling the big data analysis in the smart grid," in *Power & Energy Society General Meeting*, 2015, pp. 1–5.

[9] J. Baek, Q. H. Vu, J. K. Liu, X. y. Huang, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart grid," *IEEE Transactions on Cloud Computing*, vol. 3, no. 2, pp. 233–244, Apr./Jun. 2015.

[10] Z. Asad and M. A. R. Chaudhry, "A two-way street: Green big data processing for a greener smart grid," *IEEE Systems Journal*, vol. 11, no. 2, pp. 784–795, Jun. 2017.

[11] A. A. Munshi and Y. A. Mohamed, "Cloud-based visual analytics for smart grids big data," in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2016, pp. 1–5.

[12] Y. Gu, H. G. Jiang, Y. C. Zhang, J. J. Zhang, T. L. Gao, and E. Muljadi, "Knowledge discovery for smart grid operation, control, and situation awarenessa big data visualization platform," in *North American Power Symposium (NAPS), 2016*, 2016, pp. 1–6.

[13] D. Koller, N. Friedman, L. Getoor, and B. Taskar, "Graphical models in a nutshell," in *Introduction to Statistical Relational Learning*, L. Getoor and B. Taskar, Eds. Cambridge, MA, USA: MIT Press, 2007.

[14] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[15] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge, UK: Cambridge University Press, 2012.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013.

[17] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search." *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[18] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. T. Chen, T. Lillicrap, H. Fan, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, Oct. 2017.

[19] R. Aggarwal and Y. H. Song, "Artificial neural networks in power systems. I. general introduction to neural computing," *Power Engineering Journal*, vol. 11, no. 3, pp. 129–134, Jun. 1997.

[20] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in *Advances in Neural Information Processing Systems*, 2002, pp. 841–848.
[21] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2001.
[22] A. J. R. Reis and A. P. A. da Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 189–198, Feb. 2005.
[23] D. Granados-Lieberman, R. J. Romero-Troncoso, R. A. Osornio-Rios, A. Garcia-Perez, and E. Cabal-Yepez, "Techniques and methodologies for power quality analysis and disturbances classification in power systems: a review," *Iet Generation Transmission & Distribution*, vol. 5, no. 4, pp. 519–529, Apr. 2011.
[24] K. J. Chen, C. W. Huang, and J. L. He, "Fault detection, classification and location for transmission lines and distribution systems: a review on the methods," *High Voltage*, vol. 1, no. 1, pp. 25–33, Apr. 2016.
[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[26] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
[27] K. M. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," arXiv preprint arXiv:1703.06870, 2017.
[28] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[29] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.
[30] Y. H. Wu, M. Schuster, Z. F. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. B. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 2016.
[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[32] K. J. Chen, J. Hu, and J. L. He, "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," *IEEE Transactions on Smart Grid*, 2017, doi: 10.1109/TSG.2016.2598881, to be published.
[33] B. J. Claessens, P. Vrancx, and F. Ruelens, "Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control," *IEEE Transactions on Smart Grid*, 2016, doi: 10.1109/TSG.2016.2629450, to be published.
[34] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine, and C. H. Kim, "Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system," in *International Conference on Intelligent Systems Applications To Power Systems*, 2007, pp. 1–6.
[35] W. C. Kong, Z. Y. Dong, Y. W. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, 2017, doi: 10.1109/TSG.2017.2753802, to be publihsed.
[36] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015, pp. 55–64.
[37] W. Z. Wu, K. J. Chen, Y. Qiao, and Z. X. Lu, "Probabilistic short-term wind power forecasting based on deep neural networks," in *International Conference on Probabilistic Methods Applied To Power Systems*, 2016, pp. 1–8.
[38] W. He, "Load forecasting via deep neural networks," *Procedia Computer Science*, vol. 122, pp. 308–314, 2017.
[39] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.
[40] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," arXiv preprint arXiv:1312.6203, 2013.
[41] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proceedings of the 30th Annual International Conference on Neural Information Processing Systems*, Dec. 2016, pp. 3844–3852.
[42] H. B. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[43] B. L. Peng, Z. D. Lu, H. Li, and K. F. Wong, "Towards neural network-based reasoning," arXiv preprint arXiv:1508.05508, 2015.
[44] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," in *Proceedings of the 29th Annual International Conference on Neural Information Processing Systems*, vol. 1, Dec. 2015, pp. 1693–1701.
[45] A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwińska, S. G. Colmenarejo, E. Grefenstette, T. Ramalho, J. Agapiou, A. P. Badia, K. M. Hermann, Y. Zwols, G. Ostrovski, A. Cain, H. King, C. Summerfield, P. Blunsom, K. Kavukcuoglu, and D. Hassabis, "Hybrid computing using a neural network with dynamic external memory," *Nature*, vol. 538, no. 7626, pp. 471–476, Oct. 2016.
[46] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
[47] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68–80, Jun. 2012.
[48] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[49] K. J. Chen, J. Hu, and J. L. He, "A framework for automatically extracting overvoltage features based on sparse autoencoder," *IEEE Transactions on Smart Grid*, 2016, doi: 10.1109/TSG.2016.2558200, to be published.
[50] Y. Wang, Q. X. Chen, C. Q. Kang, Q. Xia, and M. Luo, "Sparse and redundant representation-based smart meter data compression and pattern extraction," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2142–2151, May 2017.
[51] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," arXiv preprint arXiv:1312.6114, 2013.
[52] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and variational inference in deep latent Gaussian models," arXiv preprint arXiv:1401.4082, 2014.
[53] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 28th Annual International Conference on Neural Information Processing Systems*, vol. 2, Dec. 2014, pp. 2672–2680.
[54] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proceedings of the 30th Annual International Conference on Neural Information Processing Systems*, Dec. 2016, pp. 2234–2242.
[55] Y. Z. Chen, Y. S. Wang, D. Kirschen, and B. S. Zhang, "Model-free renewable scenario generation using generative adversarial networks," arXiv preprint arXiv:1707.09676, 2017.
[56] Z. H. Zhou, *Machine Learning*. Beijing, China: Tsinghua University Press, 2016.
[57] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, May 1992.
[58] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, no. 1, pp. 237–285, Jan. 1996.
[59] M. Glavic, R. Fonteneau, and D. Ernst, "Reinforcement learning for electric power system decision and control: Past considerations and perspectives," in *Proceedings of the 20th IFAC World Congress*, 2017.
[60] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *1st IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2010, pp. 409–414.
[61] Z. Wen, D. ONeill, and H. Maei, "Optimal demand response using device-based reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2312–2324, Sep. 2015.
[62] B.-G. Kim, Y. Zhang, M. van der Schaar, and J.-W. Lee, "Dynamic pricing and energy consumption scheduling with reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2187–2198, Sep. 2016.
[63] D. Ernst, P. Geurts, and L. Wehenkel, "Tree-based batch mode reinforcement learning," *The Journal of Machine Learning Research*, vol. 6, pp. 503–556, Dec. 2005.

[64] F. Ruelens, B. J. Claessens, S. Vandael, B. De Schutter, R. Babuška, and R. Belmans, "Residential demand response of thermostatically controlled loads using batch reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2149–2159, Sep. 207.

[65] Q. Q. Sun, W. H. Yu, N. Kochurov, Q. Hao, and F. Hu, "A multi-agent-based intelligent sensor and actuator network design for smart house and home automation," *Journal of Sensor and Actuator Networks*, vol. 2, no. 3, pp. 557–588, Aug. 2013.

[66] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: Reinforcement learning framework," *IEEE transactions on Power Systems*, vol. 19, no. 1, pp. 427–435, Feb. 2004.

[67] J. G. Vlachogiannis and N. D. Hatziargyriou, "Reinforcement learning for reactive power control," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1317–1325, Aug. 2004.

[68] D. Ernst, M. Glavic, F. Capitanescu, and L. Wehenkel, "Reinforcement learning versus model predictive control: a comparison on a power system problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 517–529, 2009.

[69] F. Daneshfar and H. Bevrani, "Load–frequency control: a GA-based multi-agent reinforcement learning," *IET Generation, Transmission & Distribution*, vol. 4, no. 1, pp. 13–26, Mar. 2010.

[70] R. Hadidi and B. Jeyasurya, "Reinforcement learning based real-time wide-area stabilizing control agents to enhance power system stability," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 489–497, 2013.

[71] H. W. Wang, T. W. Huang, X. F. Liao, H. Abu-Rub, and G. Chen, "Reinforcement learning for constrained energy trading games with incomplete information," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3404–3416, Oct. 2017.

[72] H. W. Wang, T. W. Huang, X. F. Liao, H. Abu-Rub, and G. Chen, "Reinforcement learning in energy trading game among smart microgrids," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 8, pp. 5109–5119, Aug. 2016.

[73] S. Vandael, B. Claessens, D. Ernst, T. Holvoet, and G. Deconinck, "Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1795–1805, Jul. 2015.

[74] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971, 2015.

[75] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proceedings of the 33rd International Conference on Machine Learning*, 2016, pp. 1928–1937.

[76] P. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," arXiv preprint arXiv:1706.03741, 2017.

[77] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Proceedings of the 13th Annual International Conference on Neural Information Processing Systems*, 2000, pp. 1008–1014.

[78] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.

[79] Y. S. Abumostafa, M. Magdonismail, and H. T. Lin, *Learning from Data: A Short Course*. Amlbook, 2012.

[80] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Transactions on Smart Grid*, 2017, doi: 10.1109/TSG.2017.2681962, to be published.

[81] A. Ng, *Machine Learning Yearning*, 2017, to be published.

[82] Z. H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[83] X. H. Liang, X. Li, R. X. Lu, X. D. Lin, and X. M. Shen, "UDP: Usage-based dynamic pricing with privacy preservation for smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 141–150, Mar. 2013.

[84] M. Wen, R. X. Lu, K. Zhang, J. S. Lei, X. H. Liang, and X. M. Shen, "PaRQ: A privacy-preserving range query scheme over encrypted metering data for smart grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 178–191, Jun. 2013.

[85] C. I. Fan, S. Y. Huang, and Y. L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 666–675, Feb. 2014.

[86] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proceedings of the 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2006, pp. 486–503.

[87] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the 3rd Theory of Cryptography Conference*, 2006, pp. 265–284.

[88] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *54th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2013, pp. 429–438.

[89] J. Duchi, M. J. Wainwright, and M. I. Jordan, "Local privacy and minimax bounds: Sharp rates for probability estimation," in *Proceedings of the 27th Annual International Conference on Neural Information Processing Systems*, 2013, pp. 1529–1537.

[90] A. D. Sarwate and K. Chaudhuri, "Signal processing and machine learning with differential privacy: Algorithms and challenges for continuous data," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 86–94, Sep. 2013.

[91] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.

[92] G. Ács and C. Castelluccia, "I have a dream!(differentially private smart metering)," in *Proceedings of the 13th International Conference on Information Hiding*, 2011, pp. 118–132.

[93] J. Zhao, T. Jung, Y. Wang, and X. Y. Li, "Achieving differential privacy of data disclosure in the smart grid," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, 2014, pp. 504–512.

[94] M. Backes and S. Meiser, "Differentially private smart metering with battery recharging," in *Proceedings of the 8th International Workshop on Data Privacy Management and Autonomous Spontaneous Security*, 2014, pp. 194–212.

[95] W. W. Jia, H. J. Zhu, Z. F. Cao, X. L. Dong, and C. X. Xiao, "Human-factor-aware privacy-preserving aggregation in smart grid," *IEEE Systems Journal*, vol. 8, no. 2, pp. 598–607, Jun. 2014.

[96] A. Yassine, A. A. Nazari Shirehjini, and S. Shirmohammadi, "Smart meters big data: Game theoretic model for fair data sharing in deregulated smart grids," *IEEE Access*, vol. 3, pp. 2743–2754, Dec. 2015.

[97] Z. Erkin, J. R. Troncoso-Pastoriza, R. L. Lagendijk, and F. Perez-Gonzalez, "Privacy-preserving data aggregation in smart metering systems: An overview," *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 75–86, Mar. 2013.

[98] F. J. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *1st IEEE International Conference on Smart Grid Communications*, 2010, pp. 327–332.

[99] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, 2010, pp. 61–66.

[100] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 998–1010, Jan. 2012.

[101] P. Pinson, "Introducing distributed learning approaches in wind power forecasting," in *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016, pp. 1–6.

[102] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[103] Q. H. Hu, R. J. Zhang, and Y. C. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85, pp. 83–95, Jan. 2016.

[104] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in *Proceedings of the 24th International Conference on Machine Learning*, 2007, pp. 759–766.

[105] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proceedings of the 28th Annual International Conference on Neural Information Processing Systems*, 2014, pp. 3320–3328.

[106] G. Marcus, "Deep learning: A critical appraisal," arXiv preprint arXiv:1801.00631, 2018.

[107] J. Pearl, "Theoretical impediments to machine learning with seven sparks from the causal revolution," arXiv preprint arXiv:1801.04016, 2018.

[108] J. Pearl and E. Bareinboim, "Transportability of causal and statistical relations: A formal approach," in *Proceedings of the 11th International Conference on Data Mining Workshops*, 2012, pp. 540–547.

[109] E. Bareinboim and J. Pearl, "Transportability of causal effects: completeness results," in *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, 2012, pp. 698–704.

[110] J. Pearl, "The *do*-calculus revisited," arXiv preprint arXiv:1210.4852, 2012.
[111] S. Athey and G. Imbens, "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 27, pp. 7353–7360, Jul. 2016.
[112] S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 2017, doi: 10.1080/01621459.2017.1319839, to be published.
[113] F. Johansson, U. Shalit, and D. Sontag, "Learning representations for counterfactual inference," in *Proceedings of the 33rd International Conference on Machine Learning*, 2016, pp. 3020–3029.
[114] S. Rusitschka, K. Eger, and C. Gerdes, "Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain," in *1st IEEE International Conference on Smart Grid Communications*, 2010, pp. 483–488.
[115] F. Ye, Y. Qian, and R. Q. Hu, "An identity-based security scheme for a big data driven cloud computing framework in smart grid," in *IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.
[116] H. Karimipour and V. Dinavahi, "Accelerated parallel WLS state estimation for large-scale power systems on GPU," in *North American Power Symposium*, 2013, pp. 1–6.
[117] M. Annunziata and G. Bell, "Powering the future: Leading the digital transformation of the power industry," General Electric, Tech. Rep., 2015.
[118] OMNETRIC Group. (2015) Integrated smart grid solutions for the digital era. [Online]. Available: http://omnetricgroup.com/modules/frontend/downloads/17_BR_Integrated_SmartGrid.pdf
[119] P. C. Evans and M. Annunziata, "Industrial internet: Pushing the boundaries of minds and machines," General Electric, Tech. Rep., 2012.
[120] AT&T NEWSROOM. (2015, Feb.) At&t and ge strengthen alliance, fuel innovation with smart energy solutions. [Online]. Available: http://about.att.com/story/att_and_ge_alliance_leads_to_smart_energy_solutions.html
[121] M. Egan. (2016, Jul.) GE and Microsoft join forces in the industrial data cloud. GE Reports. [Online]. Available: http://www.gereports.com/ge-microsoft-join-forces-give-industrial-cloud-power/
[122] OMNETRIC Group. (2015) Achieving quality and speed in analytics with data discovery. [Online]. Available: http://omnetricgroup.com/modules/frontend/downloads/15_BR_Data_Discovery.pdf

Kunjin Chen (S'17) received the B.Sc. degree in electrical engineering from Tsinghua University, Beijing, China, in 2015. Currently, he is a Ph.D. student with the Department of Electrical Engineering.

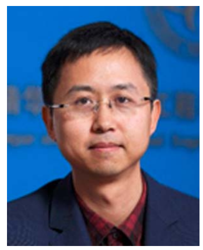
His research interests include applications of machine learning and data science in power systems.

Ziyu He recieved his B.E. degree from Zhejiang University in 2015 and his M. S. degree from Columbia University in 2017. He is currently a Ph.D. student in the department of industrial and systems engineering at University of Southern California.

His research interests are optimization, machine learning and their applications in energy.

Shan X. Wang (M'88–SM'06–F'09) received the B.S. degree in physics from the University of Science and Technology of China, Hefei, China, in 1986, the M.S. degree in physics from Iowa State University, Ames, IA, USA, in 1988, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1993.

He currently serves as the Director of the Stanford Center for Magnetic Nanotechnology and as a Professor of Materials Science and Engineering, jointly of Electrical Engineering, and by courtesy, a Professor of Radiology at Stanford University, Stanford, CA, USA. He is a CoPrincipal Investigator of the Stanford-led Center for Cancer Nanotechnology Excellence and Translation, Bethesda, MD, USA. His research interests lie in magnetic nanotechnologies and information storage in general and include magnetic biochips, in vitro diagnostics, cell sorting, magnetic nanoparticles, nanopatterning, spin electronic materials and sensors, magnetic inductive heads, as well as magnetic-integrated inductors and transformers.



Jun Hu (M'10) received his B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Department of Electrical Engineering, Tsinghua University in Beijing, China, in July 1998, July 2000, July 2008.

Currently, he is an associate professor in the same department. His research fields include overvoltage analysis in power system, sensors and big data, dielectric materials and surge arrester technology.

Licheng Li received the B.E.E degree from Tsinghua University, Beijing, China, in 1963.

He is a Member of the Chinese Academy of Engineering. He is currently a Professor at the South China University of Technology, Guangdong, China. He is also with the Electric Power Research Institute, China Southern Power Grid Company Ltd., Guangdong. His main research interests include HVDC, voltage-source converter-HVDC technology, high-voltage engineering, and power system analysis.

Jinliang He (M'02–SM'02–F'08) received the B.Sc. degree from Wuhan University of Hydraulic and Electrical Engineering, Wuhan, China, the M.Sc. degree from Chongqing University, Chongqing, China, and the Ph.D. degree from Tsinghua University, Beijing, China, all in electrical engineering, in 1988, 1991 and 1994, respectively.

He became a Lecturer in 1994, and an Associate Professor in 1996, with the Department of Electrical Engineering, Tsinghua University. From 1997 to 1998, he was a Visiting Scientist with Korea Electrotechnology Research Institute, Changwon, South Korea, involved in research on metal oxide varistors and high voltage polymeric metal oxide surge arresters. From 2014 to 2015, he was a Visiting Professor with the Department of Electrical Engineering, Stanford University, Palo Alto, CA, USA. In 2001, he was promoted to a Professor with Tsinghua University. He is currently the Chair with High Voltage Research Institute, Tsinghua University. He has authored five books and 400 technical papers. His research interests include overvoltages and EMC in power systems and electronic systems, lightning protection, grounding technology, power apparatus, and dielectric material.