Imitation and Transfer Q-Learning-Based Parameter Identification for Composite Load Modeling

Jian Xie, *Student Member, IEEE*, Zixiao Ma, *Graduate Student Member, IEEE*, Kaveh Dehghanpour, *Member, IEEE*, Zhaoyu Wang, *Member, IEEE*, Yishen Wang, *Member, IEEE*,
AQ1 Ruisheng Diao, *Senior Member, IEEE*, and Di Shi, *Senior Member, IEEE*

***Abstract*—Fast and accurate load parameter identification has a large impact on power systems operation and stability analysis. This article proposes a novel Imitation and Transfer Q-learning (ITQ)-based method to identify parameters of composite constant impedance-current-power (ZIP) and induction motor (IM) load models. Firstly, an imitation learning process is introduced to improve the exploitation and exploration processes. Then, a transfer learning method is employed to overcome the challenge of time-consuming optimization when dealing with new identification tasks. An associative memory is designed to realize dimension reduction, knowledge learning and transfer between different identification tasks. Agents can exploit the optimal knowledge from source tasks to accelerate the search rate in new tasks and improve solution accuracy. A greedy action selection rule is adopted for agents to balance the global and local search. The performance of the proposed ITQ approach has been validated on a 68-bus test system. Simulation results in multi-test cases verify that the proposed method is robust and can estimate load parameters accurately. Comparisons with other methods show that the proposed method has superior convergence rate and stability.**

***Index Terms*—Load modeling, parameter identification, transfer learning, reinforcement learning, imitation learning.**

I. Introduction

As AN important part of power system analysis, electrical load modeling has a critical impact on the stable operation of power grids [1]–[3]. Incorrect load models may lead to completely biased results for system operation status and stability evaluation [4]–[7]. Due to time-variability, complex composition and non-linearity, fast and accurate load modeling still remains a challenging problem. Therefore, it is imperative to identify load model parameters accurately and rapidly to help provide more reliable results for real-time power system operation.

AQ2 Manuscript received November 18, 2019; revised May 11, 2020 and July 20, 2020; accepted September 17, 2020. This work was supported by the Power Systems Engineering Research Center under Grant PSERC S-84G. Paper no. TSG-01746-2019. *(Corresponding author: Zhaoyu Wang.)*

Jian Xie, Zixiao Ma, Kaveh Dehghanpour, and Zhaoyu Wang are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: jianx@iastate.edu; zma@iastate.edu; kavehd@iastate.edu; wzy@iastate.edu).

Yishen Wang, Ruisheng Diao, and Di Shi are with the Department of AI and System Analytics, GEIRI North America, San Jose, CA 95134 USA (e-mail: yishen.wang@geirina.net; ruisheng.diao@geirina.net; di.shi@geirina.net).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2020.3025509

Based on load models' characteristics, conventional load models can be categorized into three types: static load models, dynamic load models and composite load models. In static load models, active and reactive power can be expressed as functions of bus voltage and frequency. Common static load models include static load model which comprised of constant impedance Z, constant current I and constant power P loads (ZIP) model [8] and exponential model [9]. Dynamic load models can represent the relationship between load active/reactive power and bus voltage. Representative dynamic loads are induction motor (IM) load and exponential recovery load model (ERL) [10]. IM load model is considered to be a physical model since it is derived from the equivalent circuit of an IM [11]. Numerous studies have shown that a single static or dynamic model cannot sufficiently replicate the dynamic behavior of the actual load. Therefore, composite load models, combining ZIP and IM have been adopted by most of the utilities to represent the actual load, which can provide more accurate characteristics [12].

Previous works have focused on measurement-based load identification and parameter estimation. Measurement-based methods can be classified into two categories: artificial neural network (ANN)-based methods and optimization-based methods. The ANN-based methods do not require any pre-defined physical load models and can update load outputs (i.e., active and reactive powers of loads) using the measurements in real-time. A deep learning-based technique was proposed in [13] to identify time-varying load parameters.

Optimization-based parameter estimation algorithms usually pre-define a load structure and then try to search for the optimal parameters to minimize the error between the actual power measurements and the estimated power responses. These methods can be divided into statistical techniques and heuristic techniques. Common statistical search techniques include least square (LS) method, maximum likelihood method and gradient-based method. In [14], a weighted LS method was utilized to estimate the parameters of a first order IM. However, LS methods are sensitive to outliers. Also, it can be difficult to determine the exact load parameters when the estimation process is performed over only a small number of replicated observations. A maximum likelihood approach was adopted in [15] to estimate load parameters. The two disadvantages of this method are that it is based on strong assumptions on the data structure and is sensitive to the choice of initial values. In [16], a gradient-based method was proposed to estimate

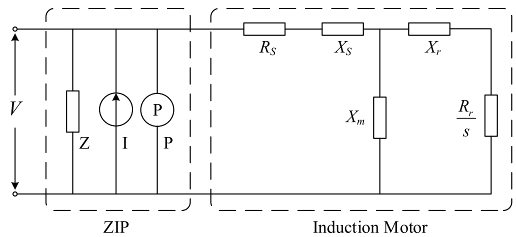


Fig. 1. Equivalent circuit of composite ZIP and IM load model [3].

parameters of a fifth-order IM load. However, gradient-based methods are sensitive to the learning rate and depend on proper initialization.

As for heuristic techniques, genetic algorithm (GA) [2] has been widely adopted to estimate the parameters of load models. GA-based methods are sensitive to the distribution of initial population of candidate solvers. In addition, premature convergence is another issue that should be considered when solutions are generated. An improved particle swarm optimization (IPSO) method has been applied in [17] to identify the unknown composite load model parameters. Unfortunately, most of the above methods are unable to exploit the prior optimization knowledge when dealing with new optimization tasks, which will result in an inefficient search when dealing with new load parameter identification tasks.

In this article, we cast the optimization problem of parameter identification for composite load model as a learning task. In applications involving non-linear optimization problems, reinforcement learning (RL) methods have been adopted to efficiently obtain optimal solutions [18]. During the RL process, agents execute actions and update their states based on designed exploration and exploitation rules. When applying RL method in power system, agents can be viewed as the candidate solutions, such as estimated load parameters; actions are used to tune the position of agents, i.e., tune the value of estimated load parameters. As an efficient RL method, Q-learning has been widely used for online optimization and control [19], [20]. However, similar to heuristic approaches, RL methods can suffer from inability to store prior agent knowledge since the initial state and action values are usually set to zero when dealing with a new optimization task, which results in time-consuming performance when identifying a large number of load parameters.

Recently, *transfer learning* has emerged as a more suitable alternative due to its ability to compensate the shortcoming of conventional RL by exploiting the prior knowledge obtained in previous time periods (i.e., source optimization tasks) [21]. This can significantly reduce the computational time for load parameter identification. In addition, imitation learning can guide a RL agent to take a more effective exploration at the initial period of RL search process and improve the exploration efficiency. Motivated by the advantages of imitation learning, a novel Imitation and Transfer Learning based Q-learning (ITQ) approach is proposed in this article, which aggregates Q-learning, transfer learning and imitation learning. The proposed method mitigates the computational burden and improves the accuracy of load parameter identification. The main contributions of this article can be summarized as follows:

- In the pre-learning stage of dealing with source optimization tasks, imitation learning is introduced to guide the RL agent to execute a more informative exploration instead of a random one.
- When dealing with a new identification task, knowledge transfer process is conducted based on the similarity between new tasks and source tasks to help a RL agent to effectively perform generalizations based on its previous experiences that are encoded within a pre-learned knowledge matrix.
- A swarm of agents are employed in the learning process to further accelerate learning rate. These interactive agents update their knowledge matrices simultaneously and share their optimal solutions during learning process.
- A greedy random search rule is developed in RL process to ensure that the proposed method can obtain high quality solutions over time.

The rest of this article is structured as follows: Section II describes the composite load model structure. Section III presents the basic principles of ITQ. The framework of ITQ based load model parameters identification is given in Section IV. Simulation results are presented in Section V, and Section VI concludes this article.

II. Composite Load Model Structure

An equivalent circuit of composite load model, consists of static ZIP and dynamic IM components connected in parallel is shown in Fig. 1. The mathematical descriptions of the active and reactive power of the ZIP component are expressed as follows:

$$P_{ZIP} = P_{ZIP,0}\left(a_p\left(\frac{V}{V_0}\right)^2 + b_p\left(\frac{V}{V_0}\right) + c_p\right) \tag{1}$$

$$Q_{ZIP} = Q_{ZIP,0}\left(a_q\left(\frac{V}{V_0}\right)^2 + b_q\left(\frac{V}{V_0}\right) + c_q\right) \tag{2}$$

where $P_{ZIP,0}$, $Q_{ZIP,0}$, V_0 are active, reactive power and root-mean-square (RMS) value of voltage in the steady state before disturbance and V is the bus voltage magnitude at a given time. In addition, ZIP parameters a_p, b_p and c_p satisfy $a_p + b_p + c_p = 1$, and a_q, b_q and c_q satisfy $a_q + b_q + c_q = 1$.

The parameters of the IM component include: stator resistance R_s, rotor resistance R_r, stator reactance X_s, and rotor reactance X_r, magnetizing reactance X_m, and the slip s.

The IM component dynamics can be expressed as follows:

$$\frac{dE'_d}{dt} = -\frac{R_r}{X_r + X_m}\left(E'_d + \frac{X_m^2}{X_r + X_m}I_q\right) - (\omega - 1)E'_q \tag{3}$$

$$\frac{dE'_q}{dt} = -\frac{R_r}{X_r + X_m}\left(E'_q - \frac{X_m^2}{X_r + X_m}I_d\right) + (\omega - 1)E'_d \tag{4}$$

$$\frac{d\omega}{dt} = -\frac{1}{2H}\left[T_0\left(A\omega^2 + B\omega + C\right) - \left(E'_d I_d + E'_q I_q\right)\right] \tag{5}$$

where H is the rotor inertia constant; A, B and C denote the torque coefficients and satisfy $A\omega^2 + B\omega + C = 1$; $\omega = 1 - s$ represents the rotation speed of the induction motor; E'_d and E'_q

refer to the d-axis and q-axis transient electromagnetic fields (EMF) of the IM. I_d and I_q are the d and q axes currents, with detailed expressions given in [3].

Given the dynamic states, parameters and bus voltage, the active and reactive power of the IM model are determined as follows:

$$P_{IM} = U_d I_d + U_q I_q \tag{6}$$

$$Q_{IM} = U_q I_d - U_d I_q \tag{7}$$

where the d-axis bus voltage U_d and the q-axis bus voltage U_q satisfy the following equation:

$$V = \sqrt{U_d^2 + U_q^2} \tag{8}$$

By aggregating the ZIP and IM active (reactive) powers, we can obtain the total active and reactive power of the composite load model [3]. In addition, another important parameter of the composite load model, is the ratio of the initial active power of the IM to the total load, which is defined as:

$$K_{pm} = \frac{P_{Im,0}}{P_0} \tag{9}$$

where P_0 denotes the initial active power of the composite load before disturbance and $P_{Im,0}$ is the initial active power of the equivalent IM.

Traditionally, the 13 parameters in equations (1)-(9) which have to be identified to fully capture the composite model, are as follows:

$$\theta = \left[R_s, X_s, X_m, X_r, R_r, H, A, B, a_p, b_p, a_q, b_q, K_{pm}\right]$$

The parameter identification process can be written as an optimization problem with the objective function of minimizing the sum of squared difference between the estimated active/reactive power and the measured active/reactive power, as:

$$\min_{\theta} h(\theta) = \frac{\sum_{k=1}^{L}\left[(P_\theta(k) - P(k))^2 + (Q_\theta(k) - Q(k))^2\right]}{L} \tag{10}$$

where L is the number of measurement samples; $P_\theta(k)$ and $P(k)$ are the estimated and measured active power; $Q_\theta(k)$ and $Q(k)$ are estimated and measured reactive power; h is the objective function representing the load model output error.

III. Basic Principles of ITQ

The overall process of implementing the ITQ is shown in Fig. 2, which includes 4 main steps: 1) RL agents learn the optimal solution for source identification tasks based on Q-learning method and store the optimal knowledge (solution) in knowledge matrix (Q-table); 2) Other agents adopt Levenberg-Marquardt algorithm (L-M) [2] to deal with the source tasks and RL agents learn from them for a more efficient search during the initial phase via imitation learning; 3) When dealing with a new load parameter identification task, defining and computing the similarities between source tasks and new task; 4) estimating the optimal knowledge matrix for the new task by exploiting the previous optimal knowledge via transfer learning.

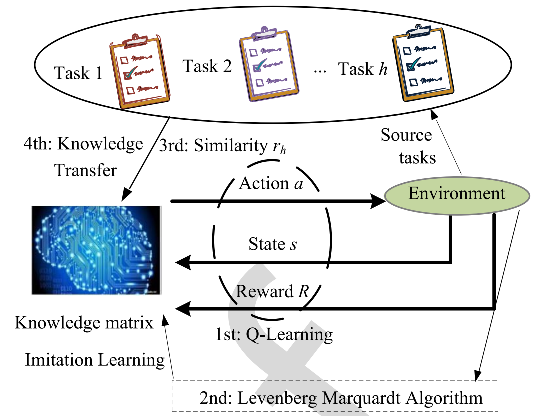


Fig. 2. Basic principle of ITQ method.

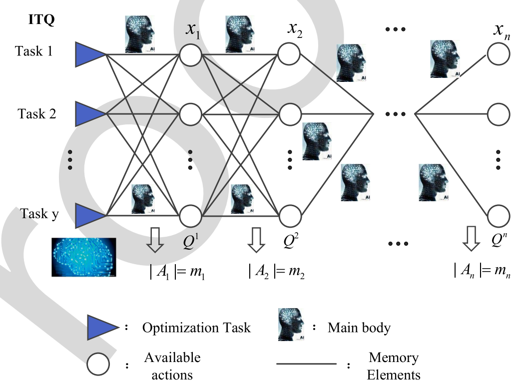


Fig. 3. Basic principle of associate memory.

A. Q-Learning

Similar to other classical RL methods, Q-learning aims to obtain an optimal policy such that a reward, R, is maximized. In the Q-learning algorithm, an agent observes the current state s and executes an action a. The system observes the corresponding results and samples a reward to the agent. The agent receives the reward and updates the Q-value corresponding to the action-state, which represents the expected estimated accumulated reward for the action-state pair. After each state transition, a new action is selected, resulting in a new state and a new reward. By continuous exploitation and exploration, the agent will eventually obtain the optimal Q-table which determines the action selection policy. In load parameter identification task, each agent can be viewed as a particle which contains estimated load parameters; actions are used to tune the value of estimated load parameter. However, there are two disadvantages for traditional Q-learning method: 1) the dimension of Q-table will increase dramatically if the number of controllable variables or the alternative actions increase; 2) using a single RL agent leads to a low knowledge learning efficiency.

However, the curse of dimensionality will emerge if the number of controllable variables grows too large in conventional Q-learning. Assuming that the number of alternative actions for a controllable variable x_i to be m_i , then the dimension of action set $|\mathbf{A}| = m_1 m_2 \cdots m_n$, where n is the number of controllable variables. If n increases significantly, the space and time complexity will increase hugely and the problem becomes intractable.

In order to avoid the curse of dimensionality, an associative memory is adopted to reduce the state-action space by decomposing the large-scale knowledge matrix (Q-table) into multiple lower-dimensional spaces [19]. As illustrated in Fig. 3, instead of adopting an extremely large-scale action set $|\mathbf{A}|$ to denote the optimization space of all the controllable variables, the multiple small-scale action sets $(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n)$ are adopted to represent the action space of each controllable variable. Consequently, each controllable variable has a corresponding memory matrix Q_i. Under such framework, the dimension of memory matrix can be greatly decreased.

Hence, each variable has a corresponding knowledge matrix. Once the action of the previous variable is determined, this action is taken as the state of the next variable, thereby forming a chain connection. By adopting the associative memory, the physical meaning of state is the same as action for load parameter identification task. A swarm of agents are adopted to improve the knowledge learning rate as there are multiple agents executing actions at the same time, which leads to simultaneous updates in Q-values of multiple state-action pairs. After introducing the swarm of agents, the ith memory matrix can be updated as:

$$\begin{cases} Q^i_{k+1}\left(s^{ij}_k, a^{ij}_k\right) = Q^i_{k+1}\left(s^{ij}_k, a^{ij}_k\right) + \alpha \Delta Q^i_k \\ \Delta Q^i_k = R^{ij}\left(s^{ij}_{k+1}, s^{ij}_k, a^{ij}_k\right) + \gamma \max_{a^i \in A_i} Q^i_k\left(s^{ij}_k, a^{ij}_k\right) \\ \qquad - Q^i_k\left(s^{ij}_k, a^{ij}_k\right) \end{cases} \tag{11}$$

where α is the learning rate; i $(i = 1, 2, \ldots, n)$ denotes the ith variable and j $(j = 1, 2, \ldots, L)$ represents the jth agent; n and L are the number of variables and agents, respectively; γ is the discount factor; subscript k denotes the iteration number; A_i denotes to the action space of agent i. ΔQ is the knowledge increment; (s_k, a_k) denotes the state-action pair at the kth iteration; $R(s_{k+1}, s_k, a_k)$ is the feedback reward of transition from state s_k to s_{k+1} after executing action a_k.

RL methods often adopt a pure strategy of greedy actions or a random global search strategy. In general, local search based on greedy strategy tends to cause the algorithm to fall into a local optimum, while random global search strategy tends to result in a long optimization time. Therefore, this article uses the ε-greedy strategy [18] to effectively balance the local search and the global search, as follows:

$$a^{ij}_{k+1} = \begin{cases} \arg\max_{a^i \in A_i} Q^i_k\left(s^{ij}_{k+1}, a^i\right), & \text{if } \varepsilon \le \varepsilon_0 \\ a_s & \text{Otherwise} \end{cases} \tag{12}$$

where ε_0 is a random number with a probability uniformly distributed in [0, 1]; ε is the exploitation rate representing the probability of a greedy action (exploitation); a_s denotes a random action (global search).

After agents execute their actions, a reward is received to evaluate corresponding state-action pair by each agent. In general, an agent will receive a larger reward if the executed action results in a better solution (i.e., smaller objective value). Hence, the reward rule is designed as follows:

$$R^{ij}\left(s^{ij}_{k+1}, s^{ij}_k, a^{ij}_k\right) = \begin{cases} \frac{1}{h^{k+1}_j}, & \text{if } h^{k+1}_j \le h^k_j \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where h^k_j is the objective function of the jth agent after the kth iteration.

B. Learning Efficiency Improvement via Imitation Learning

For a new identification task, RL agents need to execute a series of random exploitation and exploration processes to obtain the optimal policy, which consumes quite a long time without any prior knowledge and cannot meet the requirement for online load identification.

Thus, imitation learning is adopted in this section to accelerate the random search process during the initial phase of search. In the imitation process, RL agents can be regarded as *students*, which can learn and imitate other Âă*teachers* with more knowledge. In order to better guide the RL agents to update the knowledge matrix during the initial phase, a highly efficient L-M method is adopted as the teacher. The L-M algorithm is a gradient descent method. The parameter set θ updating process for L-M method is as follows:

$$\theta_{i+1} = \theta_i + \left(J^{\mathrm{T}} J + \lambda I\right)^{-1} J^{\mathrm{T}} h(\theta_i) \tag{14}$$

where θ_i denotes the estimated parameter set in the ith iteration step; J is the Jacobian matrix which can be obtained by calculating the first-order partial derivatives of estimated outputs to each parameter; λ represents the step size and I is the identity matrix.

In addition, L-M is sensitive to initial conditions and may diverge outside of the defined ranges or be trapped in a local optimal solution. In order to address these issues, some agents learn knowledge from L-M to select the state-action pair and update the knowledge matrix, the other agents update knowledge based on Q-learning and ε-greedy rule shown in (12). After each iteration, the rewards of all agents are calculated, shared and sorted. The corresponding state-action pair with the largest reward is transmitted to all imitative teachers (L-M agents). In a new iteration, the agents with larger rewards execute actions based on Q-learning principle with ε-greedy policy, while other agents with smaller reward learn from L-M to select state-action pair.

C. Knowledge Transfer via Transfer Learning

Transfer learning can be applied to discover domain-invariant intrinsic features and structures underlying two different but related domains, which establishes successful transfer and re-utilization of data information across domains. ITQ agents obtain optimal knowledge matrices (Q-tables) for source parameter identification tasks (source tasks) during the pre-learning process, the prior knowledge are then exploited as the initial knowledge matrices of a new parameter identification task (new task), thereby avoiding agents' blind explorations and improving search efficiency. This transfer

TABLE I
NUMERICAL INTERVAL OF LOAD PARAMETERS

Parameter	R_s	X_s	X_m	R_r	X_r
Range	[0.02,0.2]	[0.1,0.2]	[2,3.8]	[0.01,0.1]	[0.07,0.18]
Parameter	K_{pm}	H	A	B	a_p
Range	[0.2,0.9]	[0.5,2]	[0.2,1]	[0,1]	[0.1,0.9]
Parameter	b_p	a_q	b_q		
Range	[0.1,0.9]	[0.1,0.9]	[0.1,0.9]		

process is designed as:

$$Q^0_{ni} = \sum_{e=1}^{E} r_e Q^*_{ei}, \quad i = 1, 2, \ldots, N \tag{15}$$

where Q^0_{ni} denotes the initial knowledge matrix of the ith variable in the new task; Q^*_{ei} represents the optimal knowledge matrix of the ith parameter in the eth source task; r_e represents the similarity between the new task and the eth source task and the detailed definition of similarity between two load parameter identification tasks are described in Section IV; E denotes the number of the source task.

IV. DESIGN OF ITQ FOR LOAD PARAMETER IDENTIFICATION

In this section, the detailed steps and overall procedure to apply ITQ for load parameter identification are introduced according to the principle of ITQ.

A. Action-State Design

Although load parameters of the power system vary at different times, they always change around typical values. A larger range will affect the speed and accuracy of the algorithm, while a smaller range may exclude actual values. Therefore, in addition to algorithm performance, the range of each load parameter should be pre-designed based on its typical value in real power systems. In this article, the range of the parameter to be identified is proposed based on the typical value in actual power systems and the former related research in [12], [17], [22], as shown in Table I.

In general, the standard Q-learning algorithm is based on discrete Markov processes, which cannot be directly applied to the solution of continuous variables optimization problems. The discretization method is the most direct means to solve this problem at present. Hence, the continuous variables are divided into discrete intervals to approximate the optimal solution of the original problem with sufficient accuracy. In this article, the searching space of each continuous parameter is divided into 50 parts. For example, the search space for m_i which denotes the ith parameter in θ is $[\, m_{i1} \quad m_{i2} \quad \cdots \quad m_{i50} \,]$,which is sorted in an increasing order. To associate ITQ method with load parameters identification, we can define $Id_i \in [1, 50]$ as an index for the ith load parameter. Then, state s_i can be viewed as the current index of the ith estimated load parameter, that is $s_i = Id_i$. For instance, $s_i = 3$ means current estimation of the ith parameter is the 3rd number within the 50 parts.

Then, the action of each variable (load model parameter) is defined by:

$$\mathbf{A}_i = \left\{ a_{i,1} \quad a_{i,2} \quad \cdots \quad a_{i,50} \right\} \tag{16}$$

where $\mathbf{A}_i$ denotes the ith variable's action set; $a_{i,k}$ $(k = 1, 2, \ldots, 50)$ denotes the kth action of the ith load parameter. For instance, $a_i = 5$ means the agent selects the 5th number within the 50 parts for current iteration episode. As stated in Section III, the action set of each variable is the state set of the next variable, i.e., $\mathbf{A}_i = \mathbf{S}_{i+1}$. For the first variable, the state set is equivalent to the action set.

B. Reward Function Design

According to the description in Section III, the reward of each agent can be obtained by (13) after each iteration and a smaller objective lead to a larger reward.

C. Knowledge Transfer Design

The key to determine the transfer quality is the definition of the similarity between source task and new task. From (10) we can see that the optimization task of load parameters identification is determined by the bus voltage, active and reactive power. Hence, Fréchet distance [23] is adopted to measure the similarity between bus voltage curves, active and reactive power curves in the source tasks and new task. The Fréchet distance between the two curves is the length of the shortest leash sufficient for both to traverse their separate paths, which takes into account the location and ordering of the points along the curves. This method is widely used in curve similarity analysis. Let $\mathbf{F}$ and $\mathbf{G}$ be the bus voltage curves in the source task and new task, and the length for each curve are $\mathbf{T}$ and $\mathbf{W}$. The bus voltage in the source task is given as a function of time by $\mathbf{F}(\alpha(t))$ and $\mathbf{G}(\beta(t))$, where $\alpha(t)$ and $\beta(t)$ are two increasing functions and $\alpha(0) = 0$, $\alpha(1) = \mathbf{T}$, $\beta(0) = 0$, $\beta(1) = \mathbf{W}$. Mathematically, the Fréchet distance between the two curves is defined as:

$$\delta_F(F, G) = \inf_{\alpha,\beta} \max_{t\in[0,1]} \{d(F(\alpha(t)), G(\beta(t)))\} \tag{17}$$

where d is the Euclidean distance function.

Hence, the similarity between two bus voltage curves is determined by the equation:

$$SU(F, G) = 1 - \frac{\inf_{\alpha,\beta} \max_{t\in[0,1]}\{d(F(\alpha(t)), G(\beta(t)))\}}{\sup_{\alpha,\beta} \max_{t\in[0,1]}\{d(F(\alpha(t)), G(\beta(t)))\}} \tag{18}$$

where $SU(F, G) \in [0, 1]$, a value near 1 indicates more similarity between the two curves, while a value near 0 indicates less similarity between them.

Similarly, Fréchet distance between active (reactive) power curves are noted as SP and SQ. Then, similarity between the source load parameter identification tasks and the new identification task is defined as:

$$r = 1/3(SU + SP + SQ). \tag{19}$$

D. ITQ Parameters Setting

Suitable parameters can improve the performance of ITQ and reduce the calculation time, hence, it is crucial to choose appropriate parameters based on the generic guidelines [18]:

- The learning rate α directly determines to what extent newly acquired information overrides old information. A larger α can achieve a faster convergence rate but with a higher probability of falling into the local optimal solution. Conversely, a smaller α can lead to a slower convergence rate but ensure a higher-quality solution.
- The discount factor γ determines the importance of future rewards. Since the current optimal solution of load parameters is significant, a smaller γ should be chosen.
- The exploration rate ε allows agents to explore new action with a certain probability. A larger ε drives agents to select a greedy action rather than explore a random action.

Based on the guidelines, the four parameters of ITQ for load parameter identification can be chosen by a few trial-and error experiments and are shown in Table II. Case studies in Section V verify that these values can be viewed as a general parameters for load parameter identification task.

E. Overall Procedure

The overall process to implement the approach is shown in Fig. 4, where k_{max} denotes the maximum iteration steps and $\|Q_i^{k+1} - Q_i^k\|_2$ is the Euclidean norm of Q-value differences, and ζ is the convergence coefficient. As shown in Fig. 4, the pre-learning process is firstly executed to accumulate the optimal knowledge from the source tasks, then, agents' action strategy in the new task is initialized with transfer learning, thereby accelerating the optimization process. In real power systems, dynamic measurements can be collected after disturbance which happens in chronological order. The source task is to identify load parameters after an earlier disturbance, while the new task is the identification task based on later disturbance.

V. Case Study

This section evaluates the effectiveness of the proposed approach. The estimated results from ITQ are compared with that of the whale optimization algorithm (WOA) [24], Grey wolf optimizer (GWO) [24], IPSO [17], and classical L-M method [2]. These methods are newly invented and has been verified that they outperform GA and PSO. In order to generate the fault data, dynamic simulations are conducted on the New England 68-bus test system with composite ZIP and IM loads [13]. All simulations are undertaken in MATLAB Power System Tool (PST) and the sampling rate is 100Hz. The population size and the maximum iteration step are set as 30 and 1000 for each heuristic optimization algorithm. For ITQ, the parameters are shown in Table II.

A. Simulation Model

The 68-bus test system is a reduced-order model of the New England/New York interconnected system [13]. It contains 16 generators, 68 buses and 29 loads. Each load is described as a composite load with ZIP and IM. Load parameters identification process is carried out for the load connected to bus 27.

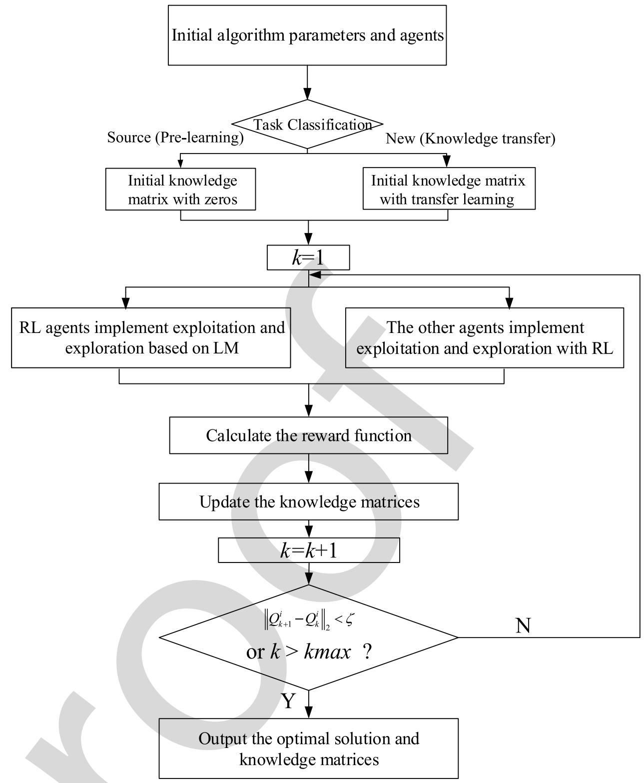


Fig. 4. Overall Procedure of ITQ.

TABLE II
PARAMETERS USED IN ITQ

Parameter	Pre-learning	Transfer learning
α	0.1	0.1
γ	0.2	0.1
ϵ	0.5	0.8

B. Pre-Learning Process

A pre-learning process needs to be firstly executed to accumulate the optimal knowledge matrices from the source tasks for ITQ algorithm. Therefore, 5 different tasks are simulated and tasks 1 and 2 are taken as source tasks. True load parameters in each task are shown in Table III. In task 1 and 3, fault occurs on the line between bus 60 and bus 61; in task 2 and 4, fault occurs on the line between bus 18 and bus 49; in task 5, fault occurs on the line between bus 19 and bus 68. The Fault type is three phase fault in all tasks.

As stated in Section III, an associative memory is designed to realize dimension reduction by decomposing the large-scale knowledge matrix (Q-table) into multiple lower-dimensional spaces. For all case studies in this article, since the searching space of each continuous parameter is divided into 50 parts, the dimensions of each low dimensional Q table is set to be 50×50.

In the pre-learning process, RL agents are initialized as zeros and a random initialization is adopted to determine the

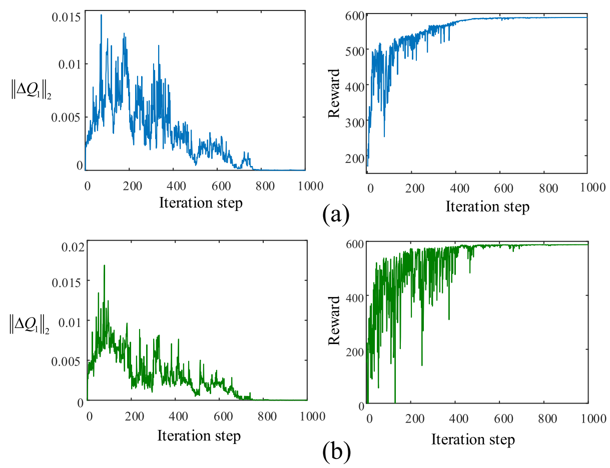


Fig. 5. Convergence of the memory matrices and reward obtained by an agent in two tasks.

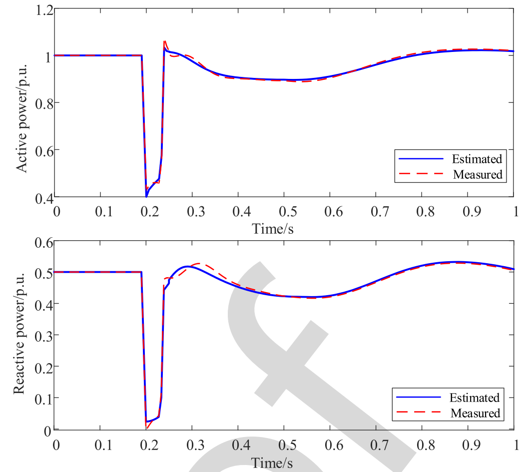


Fig. 7. Comparison between measurements and estimated outputs.

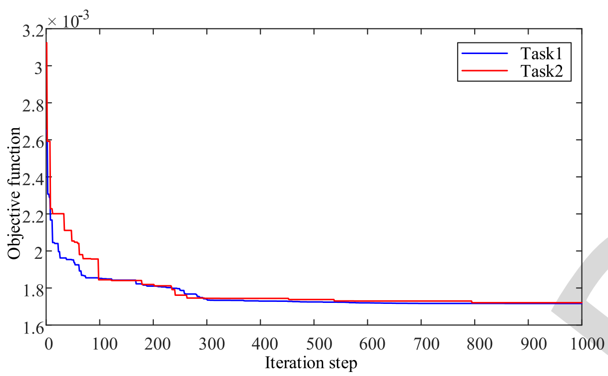


Fig. 6. Convergence of the objective functions.

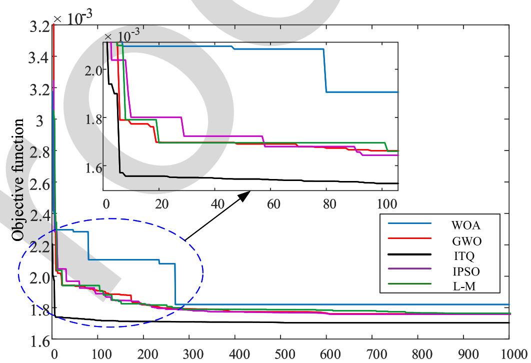


Fig. 8. Objective function obtained by five methods.

initial set of L-M agents. Fig. 5 shows the convergence curve and the reward obtained by an agent during the pre-learning process in two tasks. It is clear that each variable can converge to its own optimal knowledge matrix after 700 iteration steps. The optimal objective function during the learning process among all agents is shown in Fig. 6. It is clear that ITQ can converge to the optimal knowledge matrices for source task 1. Similarly, when applying the pre-learning process to task 2, a high quality fitness function can be obtained, as shown in Fig. 6. Fig. 7 presents the comparison between the estimated power outputs and measurements. It can be seen that the estimated outputs are very close to measurements. These results validate the highly convergence of the proposed ITQ method.

C. Transfer Learning and Comparison

With the pre-learning process completed, the optimal knowledge matrices are exploited for the online load parameters identification tasks using transfer learning. The online identification is implemented for task 3. As ITQ agents has learned the optimal knowledge from task 1 and task 2, these tasks can be viewed as source when dealing with task 3. Then, based on the definition of similarity in (19), we can compute similarities $r_{13} = 0.63$ and $r_{23} = 0.71$. Therefore, knowledge matrix for task 3 can be initialized based on (19).

Fig. 8 compares the convergence of the objective function for task 3 obtained by ITQ and other 4 algorithms, including WOA, GWO, IPSO and L-M. Reward for these optimization methods are defined as $1/h$ and h denotes the objective function. Note that all the algorithms adopt a random initialization except the proposed ITQ which is able to transfer optimal knowledge from source tasks. From Fig. 8, it is clear that ITQ can perform deep exploitation from source tasks when dealing with a new task and it can obtain the optimal solution within 150 iteration steps, which is much faster than that of the pre-learning process. The comparison verifies that the convergence rate can be dramatically accelerated by transfer learning. Compared with other methods, ITQ converge the faster and can obtain a better reward. In addition, ITQ can obtain a higher quality reward contributed to the fact that random search agents can avoid the premature convergence and search the globe optimal result. In order to further test the performance of ITQ, all the algorithms are executed with 100 runs. Fig. 9 shows the Box plots of objective functions obtained by the 5 algorithms, and it is clear that ITQ performs best and the convergence stability is higher than other algorithms.

D. Impact of Low Similarity and Limited Source Tasks

This section validates the effectiveness of ITQ with low similarity and limited source tasks. In real power systems, limited source tasks can be an obstacle for transfer learning.

TABLE III
PRE-SET PARAMETERS FOR DIFFERENT TASKS

Task	R_s	X_s	X_m	X_r	K_{pm}	R_r	a_p	a_q	b_p	b_q	H	A	B
1	0.045	0.173	2.49	0.131	0.43	0.031	0.40	0.30	0.30	0.30	1.2	0.9	0.10
2	0.113	0.104	2.21	0.081	0.71	0.045	0.55	0.25	0.15	0.35	1.1	0.5	0.83
3	0.188	0.145	3.35	0.151	0.55	0.065	0.30	0.20	0.40	0.40	0.7	0.9	0.51
4	0.151	0.112	2.83	0.163	0.62	0.021	0.61	0.15	0.23	0.42	1.4	0.7	0.29
5	0.072	0.152	3.22	0.097	0.33	0.071	0.33	0.27	0.57	0.31	0.9	0.3	0.90

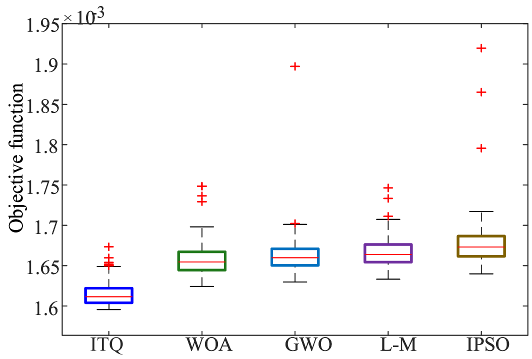


Fig. 9. Comparison of Box plot of objective function.

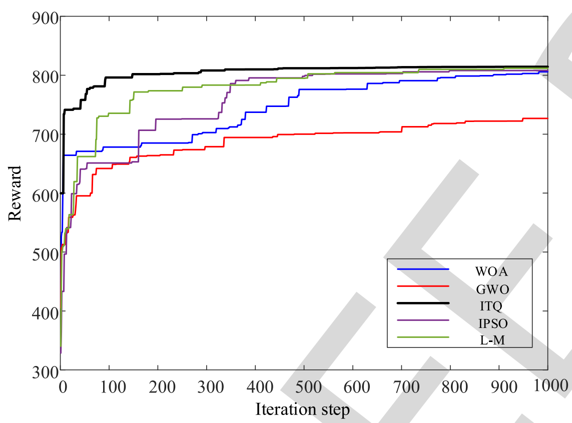


Fig. 10. Reward comparison under low similarity condition.

Similarity analysis shows that $r_{14} = 0.51$ and $r_{24} = 0.33$, and this indicates that there are few similarities between task 4 and another 2 source tasks. ITQ are adopted for task 4 to test the performance of ITQ when dealing with a new task with low similarity. Fig. 10 shows the comparison of optimization results obtained by 5 methods. It indicates that each algorithm can obtain a satisfied results and ITQ presents the biggest reward which means ITQ still has high performance even when the similarity between new task and source task is low.

Table IV presents identified results (average) from different algorithms for load parameters in task 4. For each algorithm, 150 trials have been run to obtain the optimal load parameters. For other methods, the initial set of parameters in the first trial are randomly generated and will be used for initialization in the remaining 149 trials. For ITQ, the initial knowledge matrices are the same and calculated by (15) and 19 in each trial. From the comparison results, it is clear that ITQ-based

TABLE IV
COMPARISON OF ESTIMATED PARAMETERS

Parameter	Method					
	True	ITQ	WOA	GWO	IPSO	L-M
R_s	0.151	0.1583	0.1622	0.1631	0.1628	0.1621
X_s	0.112	0.1255	0.1285	0.1311	0.1325	0.1293
X_m	2.83	2.909	3.11	3.023	3.152	3.106
X_r	0.163	0.1711	0.1832	0.1921	0.1865	0.1955
K_{pm}	0.62	0.6531	0.5885	0.6959	0.6941	0.5773
R_r	0.021	0.0358	0.0322	0.0395	0.0388	0.0331
a_p	0.61	0.5606	0.6963	0.5112	0.5232	0.6885
a_q	0.15	0.1889	0.2213	0.2515	0.2332	0.2106
b_p	0.23	0.2939	0.3121	0.2882	0.3351	0.3025
b_q	0.42	0.3101	0.2859	0.1865	0.2688	0.3232
H	1.4	1.611	1.052	1.857	1.212	1.0886
A	0.7	0.7414	0.7818	0.6543	0.6852	0.7665
B	0.29	0.3232	0.2516	0.3568	0.2158	0.3312

load parameters are closest to actual values and this is consistent with the results in Fig. 8. There are small discrepancies between estimated parameters and true values, which may be caused by the limited observability of some parameters.

E. Robustness of ITQ

Due to the complexity and nonlinearity of load models, it has been found that different load parameter combinations may lead to the same or similar dynamic response. For example, given a set of measured data (U, P and Q), multiple combinations of load model parameters may result in a same or similar reward using previous optimization methods. To test the robustness of the proposed method in searching the optimal parameters, 150 trials have been carried out for task 4 and the final reward and optimal parameters are recorded. The reward under each trial is shown in Fig. 11 and it is clear that the optimal rewards do not change much. But for other heuristic methods, rewards have large variances in 150 trials. Based on the study in [7], the eight parameters $R_s, R_r, X_r, K_{pm}, a_p, b_p, a_q, b_q$ have the highest impact on load dynamics and can be identified, while other five parameters X_s, X_m, H, A, B do not affect load dynamics and cannot be identified from voltage disturbance. Therefore, we focus on the identification results of these eight parameters.

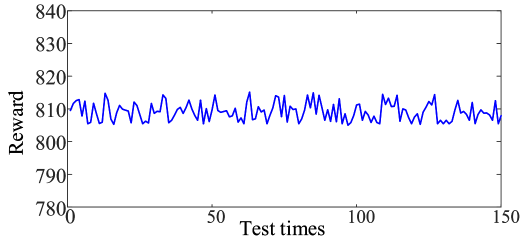


Fig. 11. Optimal reward under 150 optimization trials.

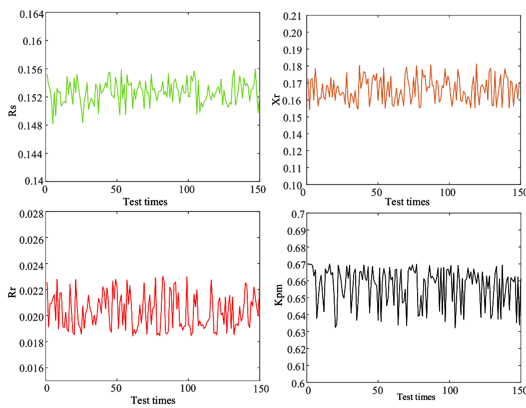


Fig. 12. Identified parameters of IM under 150 optimization trials.

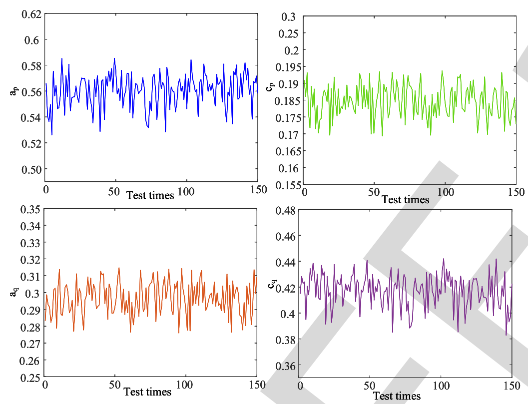


Fig. 13. Identified parameters of ZIP under 150 optimization trials.

Fig. 12 and Fig. 13 present the optimal results of these eight parameters under 150 trials. The parameters shown in these two figures are the optimal results (actions) obtained from each optimization process. It can be seen that the results of these eight parameters do not have large variances and are consistent with the corresponding true values, which corroborate the robustness of the proposed method.

In addition, as shown in Fig. 5, for a certain optimization process, the reward of the proposed ITQ method converges rapidly. In order to verify that identified parameters converge with the same rate during the RL process, Fig. 14 shows the curve of four parameters R_s, R_r, X_r, K_{pm} at each iteration step. The result is based on the data obtained in task 4 and using the proposed method. It is clear that these four parameters converge after 300 steps, which is as fast as the convergence speed of reward for task 4 shown in Fig. 10. In order to test the parameters convergence rate under each method, the

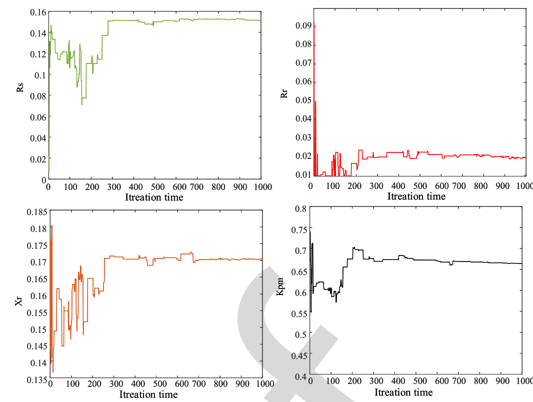


Fig. 14. Parameters converge rate.

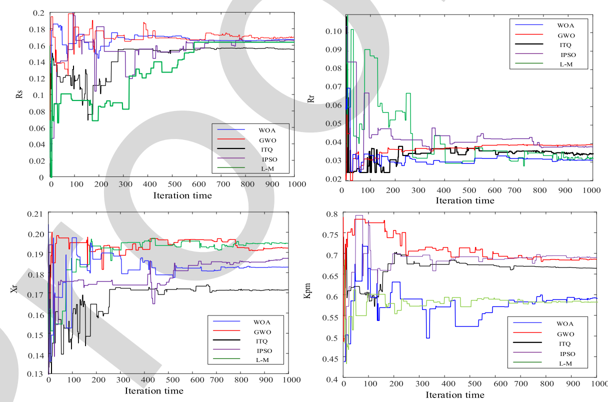


Fig. 15. Comparison of parameters converge rate.

following two figures are provided to show the comparison. Fig. 15 shows the curve of estimations for four parameters by 5 methods at each iteration step. These methods include: WOA, GWO, ITQ, IPSO and L-M. The result is based on the data obtained in task 4 which can be viewed as a new task. Fig. 16 shows statistics of the minimum step for convergence and the converge criteria requires the relative error to be smaller than 1.5% in 50 consecutive steps. The relative error in nth step σ^n is defined as:

$$\sigma^n = \left|X^{n+1} - X^n\right| / X^n \tag{20}$$

where X^n is the estimation is nth step.

From the comparisons we can see that our proposed method achieves a higher accuracy and the parameters estimated by ITQ can converge in fewer steps, which validates the effectiveness of the proposed load identification technique.

F. Computational Efficiency

In order to fully evaluate the efficiency of the proposed method, Table. V compares the computation time of optimization process of each method for task 2 and task 3, which belong to a source task and a new task, respectively. All the algorithms are implemented in MATLAB R2019a by a personal computer with Intel(R) i5 CPU at 2.6GHz with 8GB of RAM. Besides, in task 3, the proposed ITQ is able to transfer optimal knowledge from source tasks.

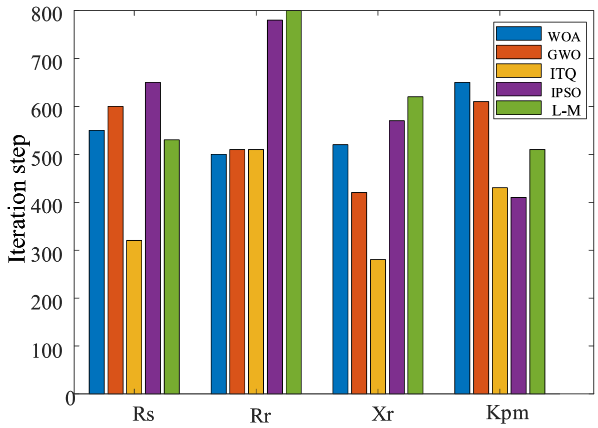


Fig. 16. Minimum iteration step to converge.

TABLE V
COMPARISON OF ESTIMATED PARAMETERS

Method	Task 2		Task 3	
	Number of iterations	Time/s	Number of iterations	Time/s
WOA	332	15.2	-	44.35
GWO	413	14.3	343	11.72
ITQ	292	19.8	8	0.58
IPSO	355	16.3	325	14.51
LM	277	17.52	371	23.92

Note that the computational time required in the optimization of composite load modeling also depends on the number of sampled data. In this study, there are 130 samples in each task. Besides, to offer a fair comparison, the same convergence criteria is used during optimization process. The criteria is for the objective function to reach a value below 1.8e-3.

From Table. V and Fig. 8, we can see that the it takes some time for ITQ agents to solve source task (pre-learning process) by greedy search and guided by teachers (L-M agents). When dealing with a new task, ITQ enables more accurate and efficient parameter identification. In power systems, power companies recorded most measurements during faults and ITQ can complete the pre-learning process by off-line learning based on these previous recorded measurements and identify load parameters in a short time when dealing with a new task.

VI. CONCLUSION AND FUTURE WORK

This article proposes an Imitation and transfer Q-learning based-based composite load parameter identification approach to accelerate the identification rate and improve the identification accuracy. An imitation learning process is introduced to improve the exploitation and exploration process of Q-learning. A transfer learning process is employed to improve the load parameter identification efficiency. Owing to the balance between greedy search and random global search rule, the proposed ITQ can avoid the premature convergence and search the global optimal result. Simulations on a 68-bus test system have validated the effectiveness of the proposed ITQ method, and the comparisons show that ITQ approach has superior convergence properties owing to the ability to exploit optimal knowledge from source tasks.

Considering the development of complex load models and time-varying load parameters, in the future work, we will extend this approach and explore up-to-date methods to identify the Western Electricity Coordinating Council (WECC) composite load model and time-varying load parameters.

REFERENCES

[1] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load modeling—A review," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5986–5999, Nov. 2018.

[2] H. Bai, P. Zhang, and V. Ajjarapu, "A novel parameter identification approach via hybrid learning for aggregate load modeling," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1145–1154, Aug. 2009.

[3] C. Wang, Z. Wang, J. Wang, and D. Zhao, "Robust time-varying parameter identification for composite load modeling," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 967–979, Jan. 2019.

[4] H. Renmu, M. Jin, and D. J. Hill, "Composite load modeling via measurement approach," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 663–672, May 2006.

[5] Z. Ma, Z. Wang, Y. Wang, R. Diao, and D. Shi, "Mathematical representation of WECC composite load model," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 5, pp. 1015–1023, Sep. 2020.

[6] Q. Huang *et al.*, "Aggregate protection response of motor loads in commercial buildings," in *Proc. IEEE/PES Transm. Distrib. Conf. Exposit. (TD)*, Denver, CO, USA, Apr. 2018, pp. 1–5.

[7] J. Ma, D. Han, R.-M. He, Z.-Y. Dong, and D. J. Hill, "Reducing identified parameters of measurement-based composite load model," *IEEE Trans. Power Syst.*, vol. 23, no. 1, pp. 76–83, Feb. 2008.

[8] P. Kundur, N. J. Balu, and M. G. Lauby, *Power System Stability and Control*, vol. 7. New York, NY, USA: McGraw-Hill, 1994.

[9] J. Zhao, Z. Wang, and J. Wang, "Robust time-varying load modeling for conservation voltage reduction assessment," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3304–3312, Jul. 2018.

[10] D. J. Hill, "Nonlinear dynamic load models with recovery for voltage stability studies," *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 166–176, Feb. 1993.

[11] K. Zhang, H. Zhu, and S. Guo, "Dependency analysis and improved parameter estimation for dynamic composite load modeling," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3287–3297, Jul. 2017.

[12] D. Han, J. Ma, R. He, and Z. Dong, "A real application of measurement-based load modeling in large-scale power grids and its validation," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1756–1764, Nov. 2009.

[13] M. Cui, M. Khodayar, C. Chen, X. Wang, Y. Zhang, and M. E. Khodayar, "Deep learning-based time-varying parameter identification for system-wide load modeling," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6102–6114, Nov. 2019.

[14] I. A. Hiskens, "Nonlinear dynamic model evaluation from disturbance measurements," *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 702–710, Nov. 2001.

[15] S. Kamoun and R. Malhamé, "Convergence characteristics of a maximum likelihood load model identification scheme," *Automatica*, vol. 28, no. 5, pp. 885–896, 1992.

[16] J. De Kock, F. Van Der Merwe, and H. Vermeulen, "Induction motor parameter estimation through an output error technique," *IEEE Trans. Energy Convers.*, vol. 9, no. 1, pp. 69–76, Mar. 1994.

[17] P. Regulski, D. S. Vilchis-Rodriguez, S. Djurovic, and V. Terzija, "Estimation of composite load model parameters using an improved particle swarm optimization method," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 553–560, Apr. 2015.

[18] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (Second Edition)* (Adaptive Computation and Machine Learning). Cambridge, MA, USA: MIT Press, 2018.

[19] X. S. Zhang, T. Yu, Z. N. Pan, B. Yang, and T. Bao, "Lifelong learning for complementary generation control of interconnected power grids with high-penetration renewables and EVs," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4097–4110, Jul. 2018.

[20] L. Yin, L. Zhao, T. Yu, and X. Zhang, "Deep forest reinforcement learning for preventive strategy considering automatic generation control in large-scale interconnected power systems," *Appl. Sci.*, vol. 8, no. 11, p. 2185, 2018.

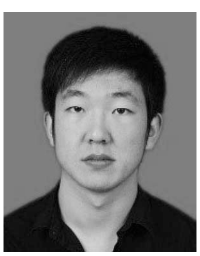
[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[22] L. Xinran, L. Peiqiang, and L. Hui, "Optimized identification strategy for composite load model parameters based on sensitivity and correlation analysis," *Trans. China Electrotechn. Soc.*, vol. 31, no. 16, pp. 181–188, 2016.

[23] T. Eiter and H. Mannila, "Computing discrete fréchet distance," Dept. Inf. Syst., Technische Universität Wien, Vienna, Austria, Rep. CD-TR 94/6, 1994.

AQ3 [24] S. Mirjalili and J.-S. Dong, *Multi-Objective Optimization Using Artificial Intelligence Techniques*. Cham, Switzerland: Springer, 2019.

Jian Xie (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. His research interests include load modeling and machine learning in power system monitoring.



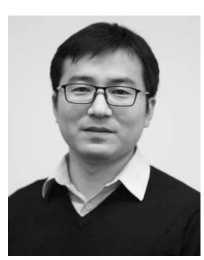
Zixiao Ma (Graduate Student Member, IEEE) received the B.S. degree in automation and the M.S. degree in control theory and control engineering from Northeastern University in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. His research interests are focused on the power system load modeling, microgrids, nonlinear control, and model reduction.

Kaveh Dehghanpour (Member, IEEE) received the B.Sc. and M.S. degrees in electrical and computer engineering from the University of Tehran in 2011 and 2013, respectively, and the Ph.D. degree in electrical engineering from Montana State University in 2017. He was a Postdoctoral Research Associate with Iowa State University. His research interests include the applications of machine learning and data-driven techniques in power system monitoring and control.

Zhaoyu Wang (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiao Tong University in 2009 and 2012, respectively, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology in 2012 and 2015, respectively. He is the Harpole-Pentair Assistant Professor with Iowa State University. His research interests include power distribution systems and microgrids, particularly on their data analytics and optimization. He is the Principal Investigator for a multitude of projects focused on these topics and funded by the National Science Foundation, the Department of Energy, National Laboratories, PSERC, and Iowa Energy Center. He is the Secretary of IEEE Power and Energy Society (PES) Award Subcommittee, the Co-Vice Chair of PES Distribution System Operation and Planning Subcommittee, and the Vice Chair of PES Task Force on Advances in Natural Disaster Mitigation Methods. He is an Editor of IEEE TRANSACTIONS ON POWER SYSTEMS, IEEE TRANSACTIONS ON SMART GRID, IEEE PES Letters, and IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY, and an Associate Editor of *IET Smart Grid*.

Yishen Wang (Member, IEEE) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2011, and the M.S. and Ph.D. degrees in electrical engineering from the University of Washington, Seattle, WA, USA, in 2013 and 2017, respectively. He is currently a Power System Research Engineer with GEIRI North America, San Jose, CA, USA. His research interests include power system economics and operation, energy storage, microgrids, load modeling, and PMU data analytics.

Ruisheng Diao (Senior Member, IEEE) received the Ph.D. degree in EE from Arizona State University, Tempe, AZ, USA, in 2009. Serving as a Project Manager, PI/co-PI and key technical contributor, he has been managing and supporting a portfolio of research projects in the area of power system modeling, dynamic simulation, online security assessment and control, dynamic state estimation, integration of renewable energy, and HPC implementation in power grid studies. He is currently with GEIRINA as a Deputy Department Head of PMU & System Analytics, in charge of several R&D projects on power grid high-fidelity simulation tools and developing new AI methods for grid operations.



Di Shi (Senior Member, IEEE) received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2007, and the M.S. and Ph.D. degrees in electrical engineering from Arizona State University, Tempe, AZ, USA, in 2009 and 2012, respectively. He currently leads the PMU & System Analytics Group, GEIRI North America, San Jose, CA, USA. His research interests include WAMS, Energy storage systems, and renewable integration. He is an Editor of IEEE TRANSACTIONS ON SMART GRID.