Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P131-cd

Deep Reinforcement Learning for Space Power Source Regulation

Tingyu Zhang

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China. E-mail: zhangtingyu@std.uestc.edu.cn

Ying Zeng

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China. E-mail: zengying2016@uestc.std.edu.cn

Yan-Feng Li

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China. E-mail: yanfengli@uestc.edu.cn

Xin Huang

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China. E-mail: xinhuang@uestc.std.edu.cn

Hong-Zhong Huang

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China. E-mail: hzhuang@uestc.edu.cn

As one of the key subsystems of space equipment, the main task of space power supply system is to ensure that it can provide continuous and stable electric energy during orbital operation as well as the bus regulation function of power supply system. For the space power supply control system represented by the S4R type, this paper proposes a dynamic analysis model of network cascade fault based on multiple charge-discharge adjustment tests which on the basis of using complex network theory to evaluate the structural reliability of the main error amplification system. Furthermore, it models the actual operation characteristics such as photovoltaic conversion and power regulation in the dynamic analysis of cascading faults, and analyzes the impact of the real-time change process on the overall reliability of the system. In this research, the regulation problem of the space power supply system is modeled as a Markov decision process model, and a power regulation algorithm based on deep reinforcement learning is further proposed to achieve intelligent monitoring and diagnosis of power supply network faults through different ways of bus regulation and filtering technology, so as to reduce the overall cascading fault risk of the space power supply distribution and supply network, while maintaining a reasonable utilization rate of stored power.

Keywords: space power system, complex network theory, power control, deep reinforcement learning.

1. Introduction

With the continuous development of space technology, spacecraft operations in space become more and more abundant. The outbreak of in-orbit service demand has promoted the development of docking technology, space debris cleaning mission to promote the birth of new tasks such as acquisition technology. Wang et al. (2021). Therefore, the space power system involved has become increasingly large, and its control system has become more complex. A performance of the proposed control strategy in both charging and discharging modes of the Battery Energy Storage Systems operating in the grid-connected mode is evaluated. Khajesalehi et al. (2015). A control architecture for autonomous operation of spacecraft power system is proposed, which is partially realized by a highly distributed software agent network. May and Loparo (2014).

Concepts for new power systems are being proposed. Okaya (2015). Researcher investigates and compares the performance of BESS models with different depths of detail. Farrokhabadi et al. (2017). The grid-connection and power control schemes of spacecraft design are constantly updated and proposed. Zhong et al. (2020). At present, real-time smart grid dispatching operation has realized the organic integration of power transmission network and signal transmission network in operation and function. Jin et al. (2018), and greatly increases the space of real-time dispatching control of power supply network. Rosato et al. (2008). Therefore, the design of an adaptive power control grid-connection algorithm can maximize the power utilization rate of the whole space power system.

Since traditional reinforcement learning algorithms could not process large continuous action state Spaces, Riedmiller first used a multilayer perceptron to approximate Q-value functions, and proposed a Neural Fitted Q Iteration algorithm. Riedmiller (2005). Lange proposed a Deep Auto-Encoder (DAE) model by combining DL model and RL method Lange and Riedmiller (2010). As a subset of machine learning, deep reinforcement learning combines the advantages of both deep learning and reinforcement learning to provide solutions to perception and decision problems of complex systems.

The interpretability of deep learning refers to the term or explanation that can be provided or briefly provided for people to understand in the application process of black box models such as deep learning. Meng et al. (2020), and some researchers interpret results by analyzing variable importance weights to identify significant features and self-attention weights to reveal persistent temporal patterns. A large number of scholars have divided the research methods on the interpretability of deep learning into post-interpretation method Kong et al. (2021), the internal explanatory power method of attention weight, and the trainable explainable deep learning model according to the pre- Ji et al. (2019), mid-Zhou et al. (2021) and post-Beven (2020) interpretation implementation in the modeling process.

In this paper, the grid-connected space power system of a certain spacecraft is taken as an example to conduct topology modeling for the grid-connected mediation scheme generated by solar cell components to overcome shadow effects during its in-orbit work. The Asynchronous Advantage Actor-Critic (A3C) parallel deep learning framework is used to conduct adaptive algorithm training for the topological network of power transmission and electrical signal communication, and the participation of the neural network in the decision-making process is interpretable.

2. Topology modeling of space power communication network

Solar array usually contains multiple cells in series to achieve a higher busbar voltage. Not only the cell in shadow lack of power generation capacity, but also will be affected by thermal effect. When a cell is completely sheltered, it will lose all photovoltaic properties, while current from the other working cells will still flow through it, what makes this battery itself does not generate voltage and cannot output power, and becomes a load and generates the heat consumption as *I2R*. The rest of the current in the battery string must generate higher voltage to compensate for the voltage loss caused by the blocked battery. Goldsmith (2023). In this case, when the shadow moves on the solar cell surface, power control unit (PCU) must cut off or connect the corresponding power module group in time, waiting for the opportunity to serve to ensure the maximum energy utilization rate and the expression of spacecraft system power generation is as Eq. (1)

$$\begin{aligned} P{=}A\times S\times\eta_1(1-k)\times z\times F_m\times F_d\times F_s \\ \times\cos\theta\times q\left[1-\left(T-298\right)K_{pt}\right] \end{aligned} \tag{1}$$

Among them, A is the area of solar array, S is the solar constant, η_1 is the average conversion efficiency of solar cells, F_m is the coefficient of canvas cloth, F_a is the end-of-life attenuation factor, F_s is the combination mismatch factor, T is the average working temperature of solar cells in the illumination area, K_{pt} is the power temperature coefficient of solar cells, k is the shielding rate of solar array, z is the comprehensive effect factor of the influence of solar array shielding on power generation, θ is the solar incident angle, q is the Kelly cosine.

Obviously, the above formula does not take into account the movement of shadows on the solar cell surface, and the grid-connected design scheme it guides has no specific response to this behavior. The power system of the space station adopts the overall scheme of multi-bus, multi-unit and multi-module grid-connected. In the state of the rail assembly, the grid-connected controller can be used to conduct grid-connected power allocation between the bus bars, so as to satisfy the safe and reliable power supply. Ma et al. (2021) . The specific structure is shown in Figure 1 and the grid-connection process of space power system is shown in Figure 2.

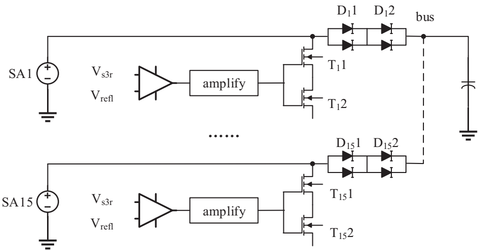

Fig. 1. Shunt regulation process of solar cell.

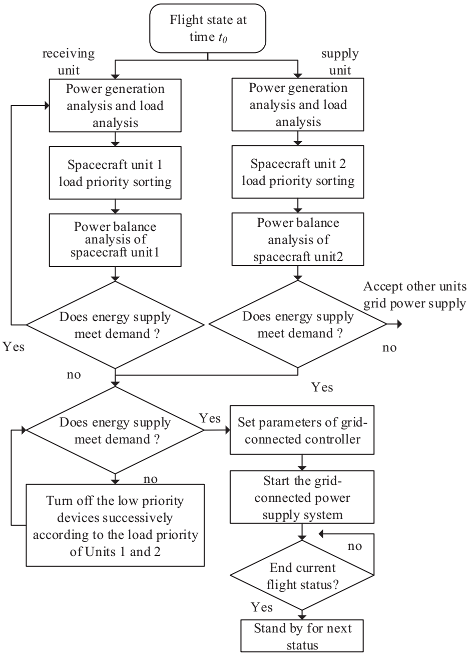
Flight state at time t_0
receiving unit
supply unit
Power generation analysis and load analysis
Power generation analysis and load analysis
Spacecraft unit 1 load priority sorting
Spacecraft unit 2 load priority sorting
Power balance analysis of spacecraft unit1
Power balance analysis of spacecraft unit2
Accept other units grid power supply
Does energy supply meet demand ?
Does energy supply meet demand ?
Yes
no
no
Yes
Does energy supply meet demand ?
Yes
Set parameters of grid-connected controller
no
Turn off the low priority devices successively according to the load priority of Units 1 and 2
Start the grid-connected power supply system
no
End current flight status?
Yes
Stand by for next status

Fig. 2. Workflow of grid connected power supply system.

The key components such as grid-connected controller and PCU are defined as nodes in the power & signal network structure. $G_E\left(V_E, E_E\right)$ is the power transmission topology, and the signal network topology is defined as $G_C\left(V_C, E_C\right)$. According to the different topologies of nodes, V_E and V_C are defined as the set of power transmission network nodes and signal transmission network nodes respectively, and N_E and N_C are the summary points of their respective networks. The specific definitions are shown in Eq.(2)- Eq.(5).

$$V_E = \left\{ v_{E,1}, v_{E,1}, \cdots v_{E,N_E} \right\} \tag{2}$$

$$V_C = \left\{ v_{C,1}, v_{C,1}, \cdots v_{C,N_C} \right\} \tag{3}$$

$$E_E = \left\{ \left(v_{E,i}, v_{E,j} \right) \middle| v_{E,i}, v_{E,j} \in V_E \right\} \tag{4}$$

$$E_C = \left\{ \left(v_{C,i}, v_{C,j} \right) \middle| v_{C,i}, v_{C,j} \in V_C \right\} \tag{5}$$

At a certain time t, the power control system needs to establish a new communication data stream

$req_t\left(v_{E,s}, v_{E,d}, bw, H\right)$ in the network. The state $s_t \in \mathscr{S}$ can be defined as Eq.(6), where $\left\{U_k^w, U_{k,j}^b\right\}$ are the alternative routing allocation scheme under the current network, as shown in Eq.(7) and Eq.(8). l_k and $l_{k,j}$ represent the maximum load of the link through which the working and switching routes are routed. $H_{E,k}^{\max}$, and $H_{C,k}^{\max}$ represent the maximum propagation parameters of the power transmission nodes coupled across the layers of the communication nodes through which the working and switching routes are routed. Zhang et al. further proposed the concept of propagation parameters under different hierarchical conditions by using the importance of edge betweenness after stratification to quantitatively evaluate the participation of each node in a certain propagation attribute (Zhang et al. 2022). It is concluded that $L\left(\mathrm{e}_{\to X}\right)$ and $L\left(\mathrm{e}_{X\to}\right)$ are the edge betweenness of node entry and exit respectively, as shown in Eq.(9). *I* and *O* are the ratio of the in-out degree of the node to the total degree of the network, n is the number of layers where the node is located, and each element in $\boldsymbol{\eta} = \left[\eta_1, \eta_2, \ldots, \eta_i\right]$ represents the contraction and expansion trend of the number of nodes in each layer.

$$s(t) = \left\{v_{E,s}, v_{E,d}, bw, H, \left\{U_k^w, U_{k,j}^b\right\}\right\} \\ \left| j = 1,2,\ldots N_J; k = 1,2,\ldots N_K \right. \tag{6}$$

$$U_k^w = \left\{l_k, H_{E,k}^{\max}, H_{C,k}^{\max}, h_k\right\} \tag{7}$$

$$U_{k,j}^b = \left\{l_{k,j}, H_{E,k}^{\max}, H_{C,k}^{\max}, h_{k,j}\right\} \tag{8}$$

$$H = \left(\|\boldsymbol{\eta}\|_2\right)^n \left[I \sum L\left(\mathrm{e}_{\to X}\right) + O \sum L\left(\mathrm{e}_{X\to}\right)\right] \tag{9}$$

3. Adaptive power transmission and signal routing optimization algorithm

The action space $\mathscr{A}$ of working routing and switching routing decision is defined as a one-dimensional vector containing a total of $N_J \cdot N_K$ routing decision actions. $a_{t,k,j} \in \mathscr{A}(t)$ allocates the *k*th alternative working routing and the *j*th alternative switching routing for the current application.

Where $P_{ss'}^a = P\left(s_{t+1} = s' \middle| s_t = s, a_t = a\right)$ is the probability of the power signal network transferring to the next state under the environmental state $s \in \mathscr{S}$. The reward function is defined as Eq.(10), which rewards the decision after the next route selection action a_t is successful. In the formula, $H_{E,DL}^{\max}\left(a_t\right)$ is the maximum criticality level of the working protection dual routing through the communication node after the establishment of the current data flow routing in the routing action, $H_{E,DL}^{\max}\left(G_C\right)$ represents the maximum criticality of the node in the signal transmission network after the establishment of the routing, *h* represents the total hops of the current selected routing, $Diam\left(G_C\right)$ is the network diameter of the signal network, and ς, η are the decision adjustment parameter of the routing.

$$\mathscr{R}\left(s_t, a_t\right) = \left[1 - \frac{H_{E,DL}\left(a_t\right)}{\eta H^{max}}\right] \cdot e^{-\varsigma \frac{H_{E,DL}^{\max}\left(a_t\right)}{H_{E,DL}^{\max}\left(G_C\right)} - \frac{h}{Diam\left(G_C\right)}} \tag{10}$$

The global optimization objective of the adaptive protection routing problem for signal transmission services is defined as shown in (11). While establishing service requests for each signal transmission data stream, the long-term optimal routing scheme for the load and risk balancing requirements of the signal transmission network in the subsequent network state continues to be sought.

The $u_{t+i,k,j}$ is a Boolean value that indicates whether the *k*th alternative working route and the corresponding *j*th switching route are taken for the data flow establishment request, and the

$\mathcal{R}\left(a_{t+i,j,k}\right)$ is an evaluation function for the rationality of the routing decision taken for the data flow establishment request. Therefore, the adaptive switching routing problem of signal transmission services studied in this paper is first established as a mixed integer linear programming form as shown below. The Eq.(11) to Eq.(13) are the constraints of the mixed integer linear programming optimization problem. Eq.(12) is to limit the sum of all established routing bandwidths on any communication line to no more than its line capacity. $y_{t+i,k,j}^{lm}$ is the number of times that the signal transmission, and $\left(v_{E,l}, v_{E,m}\right) \in E_E$ is used by the kth alternative working route and its corresponding *j*th switching route. Eq.(13) can restrict the establishment of a request $req_{t+i} \in R$ for a given data stream, and its working route does not coincide with the link used for the switching route. Eq.(14) limits the establishment of a request $req_{t+i} \in R$ for a given data stream.

$$\max \sum_{i=0}^{|R|-t} \sum_{k=1}^{N_k} \sum_{j=1}^{N_J} u_{t+i,k,j} \cdot R\left(a_{t+i,j,k}\right) \tag{11}$$

$$\sum_{i=0}^{|R|-t} \sum_{k=1}^{N_k} \sum_{j=1}^{N_J} x_{t+i,k,j} \cdot y_{t+i,k,j}^{lm} \cdot bw_{t+i} \le C_{c,lm}, \quad \forall\left(v_{E,l}, v_{E,m}\right) \in E_E \tag{12}$$

$$\sum_{i=0}^{|R|-t} \sum_{k=1}^{N_k} \sum_{j=1}^{N_J} x_{t+i,k,j} \cdot y_{t+i,k,j}^{lm} \le 1, \forall\left(v_{E,l}, v_{E,m}\right) \in E_E \tag{13}$$

$$\sum_{k=1}^{N_k} \sum_{j=1}^{N_J} x_{t+i,k,j} = 1 \tag{14}$$

To achieving the interpretability of the algorithm results and better convergence properties, this paper will use the model-free asynchronous advantage actor-critic (A3C) training framework. When the system needs to establish a new power transmission line $req_t\left(v_{E,s}, v_{E,d}, bw, H\right)$ in the network, the signal exchange process of each step is recorded one by one, and the power transmission network is extracted to generate the state vector s_t for the subsequent algorithm. Further, the policy neural network (PNN) f_{θ_π} is called to evaluate the network state, and a routing action $a_t \in \mathcal{A}(t)$ is selected in random routing strategy $\pi\left(a \middle| s_t; \theta_\pi\right)$, which is rewarded according to the feedback result and returned to the algorithm agent.

For the routing power transmission data at each moment, Eq.(15) is the long-term cumulative discount reward achieved by the routing algorithm by learning the optimal routing strategy, where $\xi \in [0,1]$ is the discount factor, which is used to adjust the attention ratio of the agent to the current reward and the long-term reward.

$$G_t = \sum_{i=0}^{\infty} \xi^i \cdot \pi\left(a \middle| s_{t+i}\right) \cdot \mathcal{R}\left(s_{t+1}, a\right) \tag{15}$$

The discounted return G_i' for each chapter is shown in Eq.(16) as an estimate of the long-term cumulative discounted return in its state. The relative advantage of the estimated results of the routing action compared to the value neural network is calculated by Eq.(17).

$$G_i' = \sum_{j=0}^{N\text{-}i} \xi^j \cdot r_{i+j} \tag{16}$$

$$A\left(\mathrm{s}_i, a_i\right) = G_i' - \upsilon(\cdot) \tag{17}$$

4. Interpretability of optimization effect

A gated recursive unit is used to control the time series dependency. Gao et al. (2020). The hidden state of the recursive component is shown in Eq.(18) - Eq.(21).

$$r_t^R = \sigma\left(x_t W_{xr} + h_{t-1} W_{hr} + b_r\right) \tag{18}$$

$$u_t^R = \sigma\left(x_t W_{xu} + h_{t-1} W_{hu} + b_u\right) \tag{19}$$

$$c_t^R = RELU\left(x_t W_{xc} + r_t e\left(h_{t-1} W_{hc}\right) + b_c\right) \tag{20}$$

$$h_t^R = \left(1-u\right) e h_{t-1} + u_t e c_t \tag{21}$$

Where e is the product of elements, σ is sigmoid function, x_t is the input of the layer at time t, W is the corresponding weight matrix, b is the bias vector, and $RELU\left(\cdot\right)$ is the activation function.

Knowledge Distillation (KD), as a teacher-student-based training model, can reduce the parameters of the model under the premise of ensuring performance. Garbay (2019). The specific steps of the process are as Table 1.There are a total of five interpretable model combinations tested, including the GRU-RS-TA-AR model that completely includes four components, as well as the RS-TA-AR model composed of three components, the GRU-TA-AT model, the GRU-RS-AR model, and the GRU-RS-TA model. The calculation formula for the degree of impact on the model prediction results is:

$$Q = \frac{\left|V_t - V_t'\right|}{V_t} \tag{22}$$

In the formula, Q represents the impact of changing a certain variable on the target value of the prediction result, which is a dimensionless value; V_t represents the root mean square error value calculated from the model prediction result, which is the target value of the model prediction result; V_t' represents the root mean square error value predicted by the model after changing a certain variable. Taking the GRU-RS-TA-AR model as an example, set the reward threshold γ to 0.6 to make statistics on the nodes involved in significant rewards or punishments and their corresponding long-term cumulative rewards. As can be seen from Table 4, if the node combinations $\left\{v_{E,33}, v_{E,47}, v_{E,48}, v_{E,79}\right\}$ and $\left\{v_{E,29}, v_{E,35}\right\}$ occur twice in succession and trigger high rewards and punishments with a relatively close interval, the long-term discount rewards associated with the node combination are retrieved from the remaining model combinations. At the same time, the predicted impact degree budget shown in Eq.(26) for G_t excluding the node is used. Taking $\left\{v_{E,29}, v_{E,35}\right\}$ as an example, dividing this node group from four Net-S models results in the corresponding long-term discount reward impact degree are shown in Figure 3.

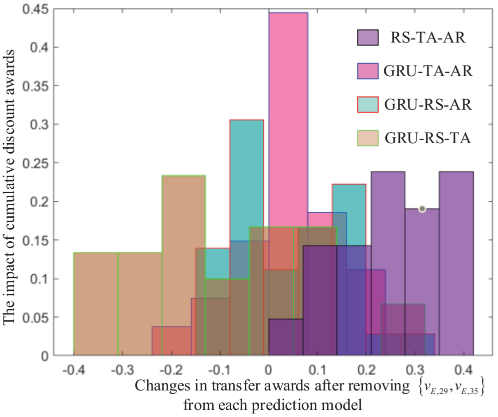

Fig. 3. Reward changes generated by different normalized prediction models for removing nodes.

Table 1. Interpretable deep learning knowledge distillation training process.

Input : Power transmission topology network state sequence
1Taking the power transmission routing model as the Net-T model
2 Take the original state sequence data as a target sequence training set.
3 The prediction target of complex model training process and training result is obtained by supplementing the training set.
4 Built different Net-S models with four interpretable components.
5. The supplementary training set and the original time series are used as training samples to train the Net-S model.
6 The error of the supplementary training set is calculated by the loss function, and the weight parameter is used to adjust the proportion of each loss function.
7 Adjust the proportion of each loss function with weight parameters.
8 Representing the key structure of Net-S model with multi-layer spatial analysis module and feature module.
9 Using supplementary training sets to train and adjust the parameters of the model to obtain a Net-S model.

10 Apply the test set to the Net-S model to predict and compare the results of the Net-T model to calculate the loss amount.

Table 2. Significant transfer reward state statistics under the GRU-RS-TA-AR model.

Nodes associated with rewards V_E, V_C	Current transfer status $s_t \to s_{t+1}$	Reward at current moment $\left|\mathcal{R}(s_{t+1},a)\right| \geq \gamma(0.6)$	Cumulative Discount Rewards G_t
……	……	……	……
$\{v_{E,21}, v_{E,24}, v_{E,29}, v_{E,35}\}$	78	0.69	52.31
$\{v_{E,33}, v_{E,47}, v_{E,48}, v_{E,79}, v_{E,80}\}$	164	0.62	61.24
……	……	……	……
$\{v_{E,33}, v_{E,47}, v_{E,48}, v_{E,49}, v_{E,52}\}$	420	-0.8	77.17
……	……	……	……
$\{v_{E,29}, v_{E,35}, v_{E,70}, v_{E,79}\}$	633	0.77	124.95

5. Conclusion

In this paper, a power transmission and communication routing algorithm based on asynchronous advantage actor-critic training framework of deep reinforcement learning is proposed to solve the problem of insufficient maneuverability of spacecraft power system during grid connection. After analyzing the interpretability of the algorithm, the main conclusions are drawn as follows: Considering the parallel characteristics of power transmission and communication dual processes involved in the model, the coupling topology network modeling of power transmission and signal transmission based on hierarchical node propagation parameters is completed. Based on Markov decision process, the real-time optimization and adaptive adjustment of time varying signal transmission network are realized by online training. Based on the explainable model components and knowledge distillation algorithm, an explainable deep learning temporal prediction model was constructed, and the model functions and input variables were further quantitatively analyzed, so as to understand the key decision basis of the model prediction process, and then the interpretation of the complex deep learning model was realized. The effect of input variable variation on prediction results under different model functions is summarized and the mechanism explanation of prediction model is realized.

Acknowledgement

This research was funded by the National Science and Technology Major Project of China under the contract No. 2017-IV-0009-0046, National Key Research and Development Program of China under the contract No. 2017YFA0701200.

References

Beven, K. (2020). Deep Learning, Hydrological Processes and the Uniqueness of Place. *Hydrological Processes 34 (16)*, 3608–13. https://doi.org/10.1002/hyp.13805.

Farrokhabadi, M, Sebastian K, Claudio C, Kankar B, and Thomas L. 2017. Battery Energy Storage System Models for Microgrid Stability Analysis and Dynamic Simulation. *IEEE Transactions on Power Systems*, 1–1. https://doi.org/10.1109/TPWRS.2017.2740163.

Gao S, Huang Y, Zhang S. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation[J]. J*ournal of Hydrology, 589*, 125188.

Garbay T, Chuquimia O and Pinna A. (2019) Distilling the knowledge in CNN for WCE screening tool[C]// *Conference on Design and Architectures for Signal and Image Processing (DASIP). Montreal：Institute of Electrical and Electronics Engineers, 2019,* 19-22.

Goldsmith, P. (2023). *Spacecraft Power Systems*. John Wiley & Sons, Inc.

Ji S, Li J, Du T. Survey on Techniques. (2019). Applications and Security of Machine Learning Interpretability[J]. *Journal of Computer Research and Development, 56(10)*, 2071-2096.

Jin, W, Yu P, Ao X, Zhang X, and Wang Y. (2018). An Approximate All-Terminal Reliability Evaluation Method for Large-Scale Smart Grid Communication *Systems. Network Operations and Management Symposium*, 1–5. https://doi.org/10.1109/NOMS.2018.8406319.

Kaelbling, L, Michael L, and Andrew M. (1996). "Reinforcement Learning: A Survey." Journal of Artificial Intelligence Research 4 (April): 237–85.

Khajesalehi J, Mohsen H, Keyhan S and Ebrahim Afjei. (2015). Modeling and Control of Quasi Z-Source Inverters for Parallel Operation of Battery Energy Storage Systems: Application to Microgrids. Electric *Power Systems Research 125*, 164–73. https://doi.org/10.1016/j.epsr.2015.04.004.

Kong X, Tang X and Wang Z. (2021). A survey of explainable artificial intelligence decision[J]. *Systems Engineering Theory & Practice, 41(02)*, 524-536.

Lange S, and Martin R. (2010). Deep Auto-Encoder Neural Networks in Reinforcement Learning. Proceedings of the International Joint Conference on Neural Networks. https://doi.org/10.1109/IJCNN.2010.5596468.

Ma J, He X and Tu B. (2021). Technical Development Achievements and Trends of Manned Spaceflight Power System in China[J]. *Aerospace Shanghai(Chinese & English), 38(03)*, 207-218.

May R, and Kenneth L. (2014). The Use of Software Agents for Autonomous Control of a DC Space Power System[C]//*12th International Energy Conversion Engineering Conference. Cleveland：International Energy Conversion Engineering*, AIAA2014-3860. https://doi.org/10.2514/6.2014-3860.

Meng, Z, Wang M, Bai J, and Hu H. (2020). Interpreting Deep Learning-Based Networking Systems. *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication. New York ： Association for Computing Machinery, 26*, 154-171. https://doi.org/10.1145/3387514.3405859.

Riedmiller M. (2005) Neural fitted Qiteration-first experiences with a data efficient neural reinforcement learning method[C]//*Machine Learning: ECML 2005. Berlin：Proceedings of the Conference on Machine Learning,* 317-328.

Okaya, S. (2015). "Advanced Concept of the Space Electric Power System Integrated with the Propulsion//*13th International Energy Conversion Engineering Conference. Orlando：International Energy Conversion Engineering, AIAA 2015*-3899.

Rosato W, and Roberto S. (2008). Modelling Interdependent Infrastructures Using Interacting Dynamical Models. *IJCIS 4 (January)*, 63–79. https://doi.org/10.1504/IJCIS.2008.016092.

Christos H. (2023). The Complexity of Markov Decision Processes. *Mathematics of Operations Researc, Volume 12,* 441-450. https://schlr.cnki.net/en/Detail/index/GARJ8099_3/SJST14120401171865.

Wang W and Yang J. (2021). Spacecraft Docking & Capture Technology: Review[J]. *Journal of Mechanical Engineering, , 57(20)*, 215-231.

Zhang T, Huang H, Li Y. (2022). Hierarchical fault propagation of command and control system[J]. *Smart Structures and Systems, 29(6)*, 791-797.

Zhong D, Tang X, Shu B and Shen B. (2020) Characteristic of Parallel Power Supply Technology for Manned Spacecraft Power System[J]. *Spacecraft engineering, 29(01)*,29-33.

Zhou Z, Cao Y, Hu C. (2020). The Interpretability of Rule-based Modeling Approach and Its Development[J]. *Acta Automatica Sinica, 47(6)*,1-19.