Optimizing Fault Detection in Electric Power Load Management Systems: A Reinforcement Learning Approach for Prioritized Test Case Selection

Meiling Feng[1,*], Ke Chen[2,*], Qiang Guo[3], Tao Meng[4], and Bin Li[5]
[1,5]School of Electric and Electronic Engineering, North China Electric Power University, Beijing, China
[2]China Electric Power Research Institute, Beijing, China
[3]State Grid Changyi Electric Power Supply Company, Changyi, China
[4]State Grid Lanzhou Power Supply Company
ml.feng@foxmail.com, 824417278@qq.com, 54447641@qq.com, mengtao@gs.sgcc.com.cn, direfish@163.com
*corresponding author

Abstract—This paper introduces a novel reinforcement learning-based method for optimizing test case selection in electric power load management systems. Addressing challenges in user-end and control loop efficiency, the study employs an Actor-Critic algorithm to prioritize test cases, enhancing detection accuracy and system stability. Through a simulated environment, test cases are represented as vectors in a multidimensional space. The model, trained over 1000 episodes, demonstrates a significant improvement in selecting optimal detection paths, as evidenced by increased total rewards and decreased loss. This approach offers a promising solution for managing complex power systems, illustrating the potential of reinforcement learning in real-world applications.

Keywords—User Side, Control Loop, Reinforcement Learning, Test Case Selection for Detection

1. INTRODUCTION

As global energy structures evolve and China's electricity system undergoes reform, the construction of new power load management systems is being steadily advanced. However, as highly intelligent systems, these new power load management systems still possess significant potential for improvement. Particularly in terms of user-end and control loop aspects, the complex and variable operations and detection stages impact the efficiency and accuracy of detection, posing challenges to the system's stability and safe operation. Consequently, there is an urgent need to develop intelligent detection technologies for the user side and control loops. Addressing the issue of load overload leading to cascading power outages and shutdowns in major components of the power system, Talaat M, Hatata A Y, and others [1] proposed an innovative, accurate, reliable, and rapid under-frequency load shedding (UFLS) technique based on the Grasshopper Optimization Algorithm (GOA). This technique aims to enhance the stability of the power system by minimizing the amount of shed load and maximally increasing the lowest swing frequency at all stages. M. A. Barik, A. Gargoom, et al. [2] developed a decentralized fault detection technique for resonant earthed distribution systems to detect single-phase-to-ground faults and identify the faulted feeder within three cycles post-occurrence. The crux of this method is the use of signals from voltage and current transformers.

Literature [3] categorizes the factors affecting the data integrity of load management terminals into two main types: terminal online rate and data collection rate. It further analyzes methods to enhance both the terminal online rate and collection rate, aiming to ensure complete transmission of metering data to the main station of the metering automation system for anomaly detection. Literature [4] mentions that the collection rate and online rate of load management terminals affect the system's capabilities in monitoring user electricity usage and load analysis. Furthermore, in load management system detection scenarios, the impact of surrounding electromagnetic interference on test results must also be considered [5].

The existing detection technologies primarily focus on system output data for fault detection, lacking research on test cases used in on-site maintenance. As a crucial tool for evaluating and maintaining power systems, these test cases play a vital role in identifying potential system issues, guiding maintenance decisions, and optimizing system performance. Accordingly, this paper proposes a prioritized selection method for test cases based on reinforcement learning. Utilizing the trial-and-error mechanism and decision optimization of reinforcement learning, the method aims to optimize the weight of detection paths. It filters an ordered subset of test cases from the full set, thereby forming an optimal detection pathway. Compared to existing methods, the proposed approach leverages the trial-and-error mechanism and decision optimization of reinforcement learning to more accurately select the test cases that yield the best detection outcomes, thereby enhancing the efficiency of maintenance and fault detection in power systems. It dynamically adjusts the selection of test cases based on the real-time state of the system and historical detection data, addressing changes in the power system environment and newly emerging fault patterns. Additionally, optimizing the detection pathway and reducing unnecessary test cases shortens the detection time and improves the accuracy and efficiency of fault diagnosis. However, the method also has some drawbacks, such as high algorithm complexity, strong dependency on data, high requirements for operational expertise, and challenges in adapting to sudden faults in real-time.

The structure of the paper is as follows. The first part is preliminary work, introducing the basics of electric power load management systems and detection cases, as well as the fundamental principles of reinforcement learning. The second part discusses the methodology, detailing the principles and processes of the proposed method. The third part presents experiments, supporting the correctness and feasibility of the method through specific experimental results. The final part

2835-2823/24/$31.00 ©2024 IEEE
DOI 10.1109/ISSSR61934.2024.00071

is a summary, providing an overarching explanation of the paper and looking towards the future.

2. PRELIMINARY

The new type of electric power load management system is designed against the backdrop of reforms in the power system. It aims at controlled and orderly electricity usage with a focus on load control and normalized load management. This system leverages communication technology, computer technology, and automation control technology to collect real-time electricity consumption data, allowing for precise load control, security assurance, and flexible interaction. The platform mainly consists of a load application module and a load control module. Its core business process involves dispatching control instructions from the management application system to the load application module, generating load control strategies, and then invoking the load control module to create load control tasks for the electricity information collection system. The "Electric Power Load Management Measures (2023 Edition)" (NDRC Operation Regulation [2023] No. 1261) explicitly states that the new electric power load management system is a software and hardware platform used for collecting, forecasting, testing, controlling, and servicing load information from electric power users, load aggregators, and virtual power plants. It serves as an information technology support system for electricity demand-side management and plays a vital role in the implementation of load management tasks.

Reinforcement Learning (Actor-Critic): Reinforcement learning is a branch of artificial intelligence that emphasizes the importance of taking appropriate actions based on the current state to maximize the value [9]. In other words, an agent makes decisions based on the current environmental state. After executing an action, it receives feedback in the form of rewards or penalties from the environment to guide better future actions. The Actor-Critic method is a reinforcement learning algorithm that combines value functions with policy. The core concept involves the Critic using the value function to evaluate the actions chosen by the Actor; the Actor adjusts its policy based on the Critic's feedback [10]. Schematic diagram of Actor Critic algorithm as shown in Figure 1.

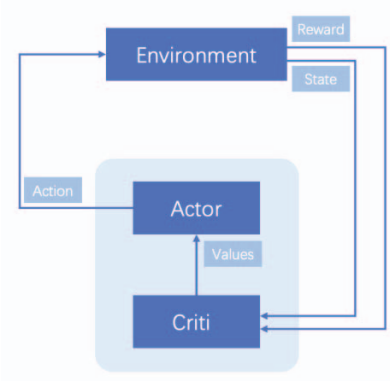

Figure 1. Schematic diagram of Actor Critic algorithm

This algorithm optimizes both the policy and the value function simultaneously, making the learning process more efficient. Therefore, the key to this algorithm lies in how to coordinate these two components to achieve better learning outcomes. The value function V(s), estimated by the Critic, represents the expected return in state s [11]. Its updated formula is as follows:

$$V_{new}(s) \leftarrow V(s) + \beta[r + \gamma V(s') - V(s)] \quad (1)$$

In this context, $V_{new}(s)$ represents the updated value function, $V(s)$ denotes the current value function, β is the learning rate of the Critic, r stands for the reward, γ is the discount factor for future rewards, and $V(s')$ is the estimated value function for the subsequent state.

The strategy of the actor is probabilistic, denoted as $\pi(s|a)$, which represents the probability of choosing action a in state s [11]. Its updated rule can be expressed as follows:

$$\theta_{new} \leftarrow \theta + \alpha[r + \gamma V(s') - V(s)]\nabla_\theta log\pi(a|s,\theta) \quad (2)$$

θ_{new} represents the updated policy parameters, θ denotes the current policy parameters, α is the learning rate of the Actor, and $\nabla_\theta log\pi(a|s,\theta)$ is the gradient of the policy, that is, the gradient of the current policy with respect to the parameters θ.

3. METHODOLOGY

In this study, we define each test case i as a vector in an n-dimensional space, denoted as $X_i = (x_1, x_2, \dots x_n)^T$. We consider a test path Q comprised of an ordered set of m such test cases, i.e., that is $Q = \{X_1, X_2, \dots X_m\}$. For this test path, we have defined a weight Y, which is used to evaluate the efficacy or quality of the path. This weight Y is calculated through a function f, namely: $Y = f(Q) = f(X_1, X_2, \dots X_m)$. The objective of this study is to select an ordered subset from the given set of test cases $\{X\}$ to form the test path Q, such that the computed path weight Y is maximized. Therefore, our optimization problem can be formally expressed as:

$$\max_{Q \subseteq \{X\}} f(Q) \quad subject\ to \quad Q = \{X_1, X_2, \dots X_m\} \quad (3)$$

Q represents an ordered subset obtained through sampling from the set {X}.

In the subsequent section, we introduce an Environment that simulates the process of test case selection. This environment encompasses a set of vectors, denoted as X_group, representing all available test cases. It maintains two crucial lists: one for tracking the vectors already selected, termed 'selected', and another for noting the vectors yet to be chosen, named 'remaining'. Upon each action execution, i.e., the selection of a vector, the environment updates these lists and computes the corresponding reward.

At each step of the operation, the environment returns the current state, which comprises information in two parts: the one-hot encoding of both the selected and the unselected vectors. We base the calculation function f on the cumulative distance between the two most recently selected vectors and design a reward mechanism accordingly.

$$reward = \sum_{i=1}^{m-1} \|X_{i+1} - X_i\| \quad (4)$$

The Actor-Critic model we adopted consists of two primary components: the Actor and the Critic. The Actor is

responsible for generating the probability distribution of actions, while the Critic evaluates the value of the current state[12]. The model optimizes the decision-making process through the collaboration of these two parts.

The training process involves multiple episodes, each representing a complete path selection procedure. In each step, the model generates an action based on the current state, and the environment updates the state and reward accordingly. Subsequently, the model computes the advantage function, a method for assessing the discrepancy between actual and expected rewards. Based on this advantage function, we calculate the losses for both the Actor and the Critic and update the model.

4. EXPERIMENTS

At this stage, since the new-type power load management system is still under construction, it is difficult to obtain relevant data. Moreover, simulated datasets can control variables, provide abundant label information, and allow repeated testing under various scenarios and conditions. Therefore, in our experiment, we initially generated a set of 100 vectors (N = 100) to characterize the test cases, with each vector comprising five dimensions (M = 5). Each dimension of the vector was derived from a normal distribution with varying means and standard deviations. These means were linearly distributed from 0 to 10, and the standard deviations linearly ranged from 1 to 2. Through this method, we constructed a multi-dimensional and variable dataset, X_group, to simulate the test case selection environment in the experiment.

In the preliminary phase of the experiment, we conducted a visual analysis of the generated dataset to ensure diversity and reasonable distribution of the data. Specifically, we plotted histograms for each dimension of the dataset to exhibit the distribution of values across different dimensions, as shown in Figure 2.

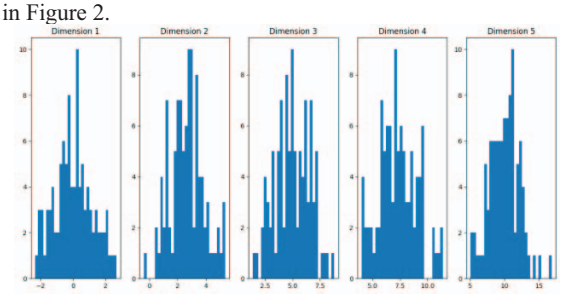

Figure 2. Dimensional distribution of the X_group dataset

To utilize the aforementioned dataset, we initialized a simulation environment, denoted as Environment. Regarding the model architecture, the input dimension was set to twice the number of vectors (input_dim = 2 * N), which accounts for the inclusion of information regarding both selected and unselected vectors. The dimension of the hidden layer was fixed at 128, and the dimension of the action space was made equivalent to the number of vectors (action_dim = N).

An Actor-Critic model was employed, and optimization was conducted using the Adam optimizer, with a learning rate set at 0.00001. This configuration aims to strike a balance between efficiency and stability in the learning process.

The training regime encompassed 1000 episodes (num_episodes=1000), with each episode representing a complete sequence of actions from the selection of the first vector to the last. During each episode, the model generated actions based on the current state of the environment and updated itself in response to feedback from the environment. We documented the total reward and loss accumulated in each episode to monitor the learning progress of the model, as illustrated in Figure 3 and Figure 4.

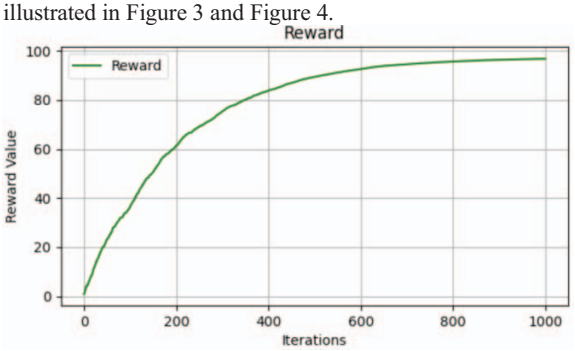

Figure 3. Total reward change chart

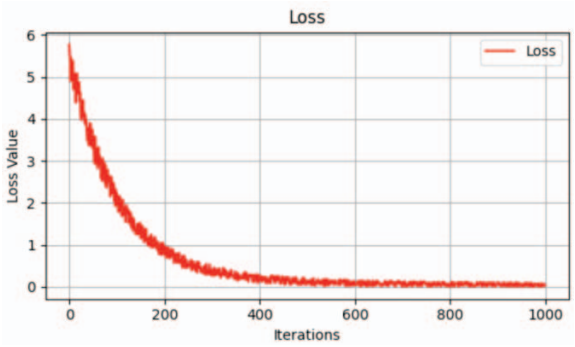

Figure 4. Loss change chart

As the number of training episodes increases, a distinct trend can be observed. The total reward gradually increases, while the loss steadily decreases. Upon the completion of 200 training episodes, the total reward reaches its peak value, and concurrently, the loss is reduced to its minimum. This phenomenon aligns with the expected outcome of the Actor-Critic algorithm.

5. CONCLUSION

This study aims to address the vector selection problem through reinforcement learning methods. The experimental results indicate that our proposed actor-critic-based use-case selection model excels in this task. After extensive rounds of training in a simulated environment, the model successfully learned how to effectively select vectors to maximize cumulative rewards. Here are the main findings and conclusions of the experiment:

Proof of Learning Capability: Experimental data shows a gradual increase in total rewards during the training process, indicating that the model effectively learned and optimized its selection strategy. Furthermore, the steady decrease in loss demonstrates the stability of the model's learning process.

Multidimensional Data Handling Capability: By using multidimensional vectors generated from normal distributions with varying means and standard deviations, we have demonstrated the model's ability to process complex and

diverse datasets. This suggests strong generalizability and adaptability of the model.

Potential Application of Reinforcement Learning in Path Selection Problems: Our methodology and experimental outcomes showcase the potential of utilizing reinforcement learning to solve complex path selection problems. This provides a valuable reference framework for future research in related fields.

Directions for Future Improvements: Although the experimental results are positive, there is room for improvement. For instance, the model's performance could be influenced by different reward mechanisms or more complex environmental structures. Future research could explore the impact of these factors on model performance and attempt to integrate more advanced reinforcement learning algorithms to further enhance efficiency and accuracy.

Overall, this study demonstrates the feasibility and effectiveness of using reinforcement learning-based methods to solve vector selection problems. Our model is not only innovative in theory but also exhibits outstanding performance in practical applications. In the future, we anticipate that this approach will be widely applied to more optimization problems, especially in scenarios that require handling large amounts of data and complex decision-making.

ACKNOWLEDGMENT

Thanks for the project supported by the Science and technology projects from State Grid Corporation（5400-202355234A-1-1-ZN）

REFERENCES

[1] Talaat M, Hatata A Y, Alsayyari A S, et al. A smart load management system based on the grasshopper optimization algorithm using the under-frequency load shedding approach[J]. Energy, 2020, 190: 116423.

[2] M. A. Barik, A. Gargoom, M. A. Mahmud, M. E. Haque, H. Al-Khalidi and A. M. Than Oo, "A Decentralized Fault Detection Technique for Detecting Single Phase to Ground Faults in Power Distribution Systems With Resonant Grounding," in IEEE Transactions on Power Delivery, vol. 33, no. 5, pp. 2462-2473, Oct. 2018, doi: 10.1109/TPWRD.2018.2799181.

[3] Xiao, Wei. "Fault Analysis and Handling Research of Power Load Management System." Dianzi Shijie [Electronic World], 2018, no. 16, pp. 145-146.

[4] Liang, Yuhui and Liu, Yang. "Research on Improving Data Integrity of Load Management Terminals for Electric Power Spot Trading Users." Jidian Xinxi [Mechanical and Electrical Information], 2019, no. 32, pp. 11-12. DOI: 10.19514/j.cnki.cn32-1628/tm.2019.32.006.

[5] Liang, Jie. "Quality Assessment of Load Management Terminal Software Based on AHP+Entropy Weight Method." Neimenggu Dianli Jishu [Inner Mongolia Electric Power Technology], 2019, vol. 37, no. 05, pp. 42-46.

[6] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," in IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26-38, Nov. 2017, doi: 10.1109/MSP.2017.2743240.

[7] Sheng WANG, Yi DING, Chengjin YE, Can WAN , Yuchang MO. Reliability evaluation of integrated electricity‐gas system utilizing network equivalent and integrated optimal power flow techniques[J].Journal of Modern Power Systems and Clean Energy, 2019, 7(06):1523-1535.

[8] National Development and Reform Commission, & National Energy Administration. (2023). Regulations on Electric Power Load Management (2023 Edition) (Order No. 2023-1261). Beijing, China: National Development and Reform Commission, National Energy Administration.

[9] G. Zhang, M. Teng, C. Chen and Z. Bie, "A Deep Reinforcement Learning Based Framework for Power System Load Frequency Control," 2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Shanghai, China, 2022, pp. 1801-1805, doi: 10.1109/ICPSAsia55496.2022.9949649.

[10] J. D. Rose, J. Mary.L, S. S, M. T. A, M. KrishnaRaj and S. Diviyasri, "Designing a Biped Robot's Gait using Reinforcement Learning's -Actor Critic Method," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 142-146, doi: 10.1109/ICICT57646.2023.10134079.

[11] R. Muduli, N. Nair, S. Kulkarni, M. Singhal, D. Jena and T. Moger, "Load Frequency Control of Two-area Power System Using an Actor-Critic Reinforcement Learning Method-based Adaptive PID Controller," 2023 IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation (SEFET), Bhubaneswar, India, 2023, pp. 1-6, doi: 10.1109/SeFeT57834.2023.10245225.

[12] R. V. J. Dayot and I. -H. Ra, "Slice Admission and Deployment Strategies in Resource-Constrained 5G Network Slices using an Actor-Critic Approach," 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS), Ise, Japan, 2022, pp. 1-4, doi: 10.1109/SCISISIS55246.2022.10001935.