Improved Precision of Hidden Hazard Identification in Power Dispatching Shift Logs Under RLHF Multi-Round Human Feedback Mechanism

Siwu Yu, Yumin He, Guobang Ban, Jintong Ma, Guanghui Xi, Lingwen Meng, Shasha Luo, Siqi Guo

Electric Power Research Institute of Guizhou Power Grid Co., Guiyang 550002, Guizhou, China

Corresponding author: Siwu Yu (e-mail: swyu2012@163.com)

ABSTRACT Reliable identification of hidden hazards in power dispatching logs is essential for the secure operation of modern electromagnetic energy systems and intelligent power infrastructures, where complex operational records and dynamically evolving risk patterns pose significant challenges. To address the insufficient identification precision caused by irregular professional terminology and continuously changing hazard characteristics in power dispatching shift logs, this study proposes a hidden hazard identification model based on multi-round Reinforcement Learning with Human Feedback (RLHF) integrated with domain knowledge. First, a power dispatching log corpus covering 35 categories of hidden hazards is established, and the Standards for Determining Major Power Safety Hazards are incorporated to construct a domain-specific knowledge base. Second, a multi-round feedback framework consisting of "pre-labeling–manual verification–model iteration" is designed, in which the Active Preference Optimization strategy is employed to improve sample utilization efficiency. Finally, the InfoRM mechanism is introduced to suppress reward hacking and enhance reward reliability. Experimental results demonstrate that the proposed model achieves precision, recall, and F1-score values of 96.82%, 95.74%, and 96.28%, respectively, outperforming the conventional BERT-BiLSTM-CRF model while exhibiting superior generalization capability under small-sample conditions. The proposed framework provides an effective solution for intelligent risk perception and operational safety assurance in electromagnetic power systems and next-generation smart grid environments.

INDEX TERMS power dispatching, hazard identification, reinforcement learning with human feedback, domain knowledge, electromagnetic power systems

I. INTRODUCTION

A. RESEARCH BACKGROUND

THE large-scale deployment of information systems in the State Grid Corporation of China has significantly enhanced power dispatching and monitoring capabilities, thereby strengthening system stability and ensuring reliable and quality energy supply to crucial sectors such as manufacturing. During the normal operation of power information systems, a large amount of log data is generated, which records a wealth of key information about system operations. On-duty logs, which document important events during power grid operation, constitute a core component of the daily work of power departments. As a form of expression for power system operation information, rapid and accurate analysis of log data plays a vital role in ensuring the safe and stable operation of power systems [1]. Traditional log processing systems are no longer capable of handling application analysis at the big data scale. In power companies, if an information system is compromised, cyberattacks and similar threats can affect the monitoring capabilities of numerous measurement and control terminal devices through various subsystems, which in turn further impacts the stable operation of relevant electrical physical equipment in the power system [2–4].

As the scale of power systems continues to expand and the complexity of managing large, dynamic industrial customers increases, power information systems have accumulated massive volumes of log data detailing operational conditions and maintenance records for sectors such as manufacturing. This data features full business coverage, diverse time types, and comprehensive time dimensions, while containing key information about power system operations [5]. In practical production work, it provides crucial guidance for operation and maintenance personnel in activities such as system

maintenance and equipment status monitoring. Real-time collection and analysis of information system logs enable the timely detection of abnormal fault information or behaviors within the system, which can reduce fault diagnosis time and business interruption duration, thereby preventing more severe consequences [6]. However, the volume of log data is often extremely large. One of the research focuses of this paper is how to identify potential safety hazards from massive intelligent on-duty log data and obtain the information required by operation and maintenance personnel.

In recent years, with the rapid development of cloud computing and big data technologies, the use of big data technologies to extract potentially useful information from massive real-time log data has attracted widespread attention from various sectors [7–9]. Currently, power information systems play an increasingly important role in power grids. Once a fault occurs, it may lead to irreversible and serious consequences. When faults occur in power information systems, they can only be detected through inefficient post-event log review, and it is impossible to identify anomalies in the initial stage and prevent them. With the emergence of technologies related to big data processing and analysis, by leveraging the massive log data stored in the system, it is possible to correlate system log data with system faults, enabling timely detection and early warning of system abnormal faults, and promptly notifying operation and maintenance personnel to take relevant measures for emergency repairs [10–12]. Therefore, intelligent hazard identification is particularly important for power log data analysis.

Reinforcement Learning from Human Feedback (RLHF) [13] is a technology that optimizes large language models by utilizing human feedback information. In RLHF, the model's behavior is no longer adjusted directly through traditional supervised learning; instead, human feedback is used to guide the model in generating outputs that better align with human preferences. RLHF enables detailed optimization of the quality, ethical compliance, and precision of content generated by the model, and is particularly suitable for complex tasks where it is difficult to clearly define a reward function [14]. By incorporating human preferences through reward models, RLHF significantly improves model performance in terms of usefulness, honesty, and harmlessness. For example, through refined preference data and RLHF training, models have made substantial progress in following complex instructions, generating safer content, and providing more satisfactory responses in specific conversational scenarios.

B. RESEARCH STATUS

In recent years, relevant research results have been achieved in the fields of risk assessment and reliability analysis for traditional networks or power information and communication networks. However, research on log analysis and early warning for power information systems remains relatively limited. This is because fault early warning resulting from the coupling and correlation between power information systems and traditional power physical systems belongs to a new research field, which requires the integration of the latest research technologies for resolution. For instance, reference [15] summarizes the new characteristics of cyber-physical coupling in risk assessment of power information and communication networks, providing new ideas for information system fault early warning; references [16,17] indicate that in the context of cyber-physical integration, the access of a large number of transmission and monitoring devices has generated a large amount of unstructured measurement data. This has significantly increased the size and complexity of sampled data, while also imposing a heavier burden on communication networks and posing greater fault risks to information systems.

After years of research and development, many excellent algorithms have been developed in the field of data mining for practical use. These mainly include K-Means, path analysis, association analysis, classification analysis, and Support Vector Machine (SVM) [18,19]. Meanwhile, applications related to massive data processing in the field of log analysis have become increasingly mature. Currently, the mainstream big data platform is still dominated by Hadoop [20]. Its architecture performs well in addressing the issue of processing massive log data and efficiently solves the pain points of traditional log systems. However, Hadoop is not well suited for real-time big data stream processing, and few real-time applications are built upon it. Therefore, it is more suitable for offline batch processing and analysis of stored historical data.

Regarding fault prediction for information system log data, the main research contents are as follows: reference [21] employs multiple machine learning methods to predict system faults respectively and then conducts a unified comparison of the execution effects of each algorithm. Reference [22] uses meta-learning to adaptively integrate multiple base predictors for fault prediction. Reference [23] adopts a dynamic rule to collect fault information and combines it with meta-learning for fault prediction. Reference [24] uses support vector machines for fault prediction. Reference [25] uses decision trees for fault prediction based on log data.

Research on RLHF has also gradually increased in recent years. This method can directly train personalized models by learning the trajectories of human preferences, avoiding the difficulty of assigning weights to personalized factors. Additionally, it can use deep neural networks to automatically learn high-level feature representations, thereby improving the model's generalization performance and learning efficiency. The RLHF framework generally consists of data collection, reward modeling, and policy optimization. In the data collection stage, it is necessary to collect a sufficient amount of human feedback data, such as user behavior data like clicks, purchases, and collections, as well as feedback data such as user ratings and comments. In the data preprocessing stage, the collected data needs to undergo preprocessing, such as removing outliers, data cleaning, and data normalization, to ensure data quality and consistency. In the model design stage, it is necessary to design a

model based on deep reinforcement learning, which can learn user preferences and behavior trajectories to generate personalized recommendation results. In the model training stage, the RLHF method is used to train the personalized recommendation model. In the model evaluation stage, the trained model needs to be evaluated to ensure its precision and reliability. Finally, in the recommendation application stage, the trained personalized model is applied to actual systems to provide users with personalized recommendation services.

The most successful applications of RLHF include ChatGPT and InstructGPT [13,26]. ChatGPT is a model used for automatic chatbots, while InstructGPT aims to execute natural language instructions. To summarize, current research on log analysis and processing in big data environments has been conducted and achieved certain results, which can provide some guidance for the work of this paper. To date, limited research has explored the application of the RLHF in hazard identification for power dispatching logs. This paper elaborates on the basic concepts and algorithm flow of applying the RLHF algorithm to power log hazard identification, and discuss the advantages, disadvantages, and experimental verification results of this method.

C. RESEARCH CONTENT AND INNOVATION POINTS

1) Research Content

(1) Construct a power dispatching log corpus containing domain knowledge and a hazard labeling system;
(2) Design a multi-round human feedback mechanism and an RLHF model architecture;
(3) Verify the model performance through comparative experiments.

2) Innovation Points

(1) Propose a domain adaptation scheme of "terminology standardization-knowledge embedding-feedback iteration";
(2) Construct a multi-round feedback framework integrating APO and InfoRM to improve sample efficiency and reward reliability;
(3) Verify the effectiveness of RLHF multi-round feedback in power hazard identification for the first time.

II. RELATED THEORETICAL FOUNDATIONS

A. CORE ISSUES IN POWER DISPATCHING HAZARD IDENTIFICATION

1) Complexity of Text Features

Power on-duty logs include elements such as equipment numbers, operation terminology, and status descriptions. These elements present three typical problems:

(1) Mixed use of professional abbreviations and aliases (e.g., "Zongjiami" refers to the vertical encryption authentication device);
(2) Ambiguous semantic boundaries (e.g., "data anomaly" requires judgment of hazard levels based on equipment types);
(3) Strong context dependence (e.g., "failure to perform supervision" needs to be associated with operation procedures).

2) Hazard Judgment Standards

According to the "Standards for Judgment of Major Power Safety Hazards", power safety hazards are divided into three major categories, including a total of 35 subcategories. Specifically, they include: equipment hazards (e.g., abnormal characteristic gas in transformers), operation hazards (e.g., unmonitored dangerous operations), and system hazards (e.g., illegal external connection of monitoring systems). Among them, 12 categories are clearly defined as major hazards, requiring immediate disposal measures.

B. RLHF TECHNICAL FRAMEWORK

The core of RLHF lies in incorporating human preference feedback. This process can be divided into three main steps: data collection, reward model training, and reinforcement learning optimization.

(1) Data Collection: In the initial stage of RLHF, it is first necessary to collect human feedback. The specific method is as follows: the model generates multiple candidate outputs based on a given input (such as a user's question or instruction), and then human evaluators assess these outputs and provide feedback. The assessment can be qualitative or quantitative, typically involving ranking or quality scoring of the outputs.
(2) Reward Modeling: Next, human feedback is converted into reward signals, which constitutes the training process of the Reward Model (RM). The reward model learns to evaluate the quality of input-output pairs through training and assigns a reward value to each output, reflecting whether the output aligns with human preferences. The goal of the reward model is to simulate the behavior of human evaluators and predict which generated texts will be considered better or more optimal in real-world applications.
(3) Reinforcement Learning Strategy Optimization: After the reward model is trained, the model optimization stage begins. This step uses reinforcement learning methods to further adjust the model's generation strategy, enabling it to produce outputs that better match human preferences given a specific input. During this process, the model's objective is to maximize the expected reward, thereby generating high-quality outputs.

III. DESIGN OF MULTI-ROUND HUMAN FEEDBACK RLHF HAZARD IDENTIFICATION MODEL

A. OVERALL ARCHITECTURE

The model architecture is divided into four layers (Figure 1), which are, from bottom to top:

(1) Data Preprocessing Layer: Implements log cleaning, terminology standardization, and feature embedding.

(2) Domain Knowledge Layer: Integrates regulatory standards and operation & maintenance experience to build a knowledge graph.
(3) RLHF Core Layer: Includes a multi-round feedback mechanism and a model optimization module.
(4) Application Output Layer: Outputs hazard types, confidence levels, and rectification suggestions.

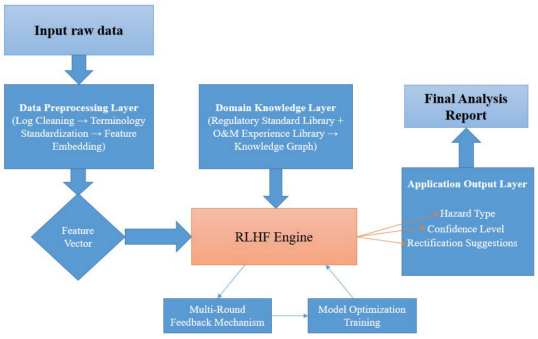

FIGURE 1. Overall Architecture

B. CORE OBJECTIVES

In the scenario of hazard identification for power dispatching on-duty logs, the core objectives of RLHF can be refined as follows:

(1) Align with the professional standards of dispatchers for hazard type determination (e.g., distinguishing between operational hazards and management hazards);
(2) Match the risk perception of operation and maintenance personnel for hazard level classification (e.g., prioritizing the identification of major hazards);
(3) Adapt to the practical needs of the dispatching scenario for trade-off between false alarms and missed alarms (e.g., allowing an extremely low false alarm rate to reduce ineffective inspections).

C. DOMAIN-ADAPTED DATA PREPROCESSING

1) Corpus Construction

A total of 120,000 on-duty logs from 2023 to 2024 were collected from a provincial-level power dispatching center. After desensitization, 8,632 hazard samples were labeled by 3 senior dispatchers. With reference to the Handbook for Safety Risk Identification and Prevention, the samples were classified by risk level into major hazards (1,245 items), general hazards (5,387 items), and non-hazards (2,000 items). Partial hazard types and examples are shown in Table 1.

TABLE 1. Main Hazard Types and Examples in Power Dispatching Logs

Major Hazard Category	Hazard Subcategory	Log Example	Risk Level
Equipment Hazards	Transformer Gas Anomaly	The acetylene content of 330 kV Xi Transformer is 0.5 μL/L, and the total hydrocarbon exceeds the standard by 30%	Major
Operational Hazards	Unsupervised Operation	110 kV Xinlingsha Line is switched to maintenance, and the supervision system is not implemented	General
System Hazards	Illegal External Connection	No encryption device is deployed at the vertical boundary of the dispatching data network	Major
Management Hazards	Abnormal Staff Status	Dispatcher Wang is listless and participates in switching operations	General

2) Terminology Standardization and Feature Embedding

Terminology Processing: Based on the indexing rules of electrical engineering vocabulary, a dictionary containing 2,386 professional terms was constructed to support terminology standardized (e.g., Kongbao was normalized to Control and Protection System);

Feature Extraction: BERT is used to extract contextual semantic features (dimension: 768), and Word2Vec is used to extract term features (dimension: 200). A joint feature vector is formed after weighted fusion via the attention mechanism.

D. DESIGN OF MULTI-ROUND HUMAN FEEDBACK MECHANISM

1) Feedback Process and Sample Selection

A three-round feedback iteration process is designed:

First Round: The Actor model initially screens high-confidence samples (confidence ≥ 0.9), and initial preference data is formed after manual verification;

Second Round: The APO strategy is used to screen the 20% of samples with the highest uncertainty. Uncertainty is quantified using entropy-based metric, defined as

$$H = -\sum_{c=1}^{n} p(c) \log p(c),$$

where $p(c)$ is the model's predicted probability for the c-th hazard category and n is the total number of hazard categories. Samples with higher entropy values indicate greater ambiguity in the model's predictions and are prioritized for expert labeling;

Third Round: For misjudged samples (e.g., confusing vertical encryption missing with horizontal isolation missing), supplementary contextual labels are introduced and prompt design is refined. The number of samples in each round of feedback increases in a 1:2:3 ratio, and the total number of samples is controlled within 2,000 to reduce labor costs.

E. REWARD MODEL OPTIMIZATION

1) Basic Reward Calculation

The Reward Model adopts a three-layer transformer architecture. It takes log features and knowledge graph association vectors as input and outputs a reward value ranging from 0 to 10. Samples that fully match hazard features get 10 points, while those missing key entities are deducted 3–5 points;

2) InfoRM Improvement

A variational information bottleneck objective function is introduced to filter out features irrelevant to human preferences. In the power dispatching scenario, reward hacking manifests as the model overfitting to superficial log patterns (e.g., repeatedly assigning high rewards to logs mentioning transformer without verifying gas parameter thresholds) rather than genuine hazard characteristics. To validate the effectiveness of InfoRM, an ablation experiment was conducted by removing the InfoRM mechanism. The results showed that the reward hacking rate (defined as the proportion of samples with inconsistent reward assignments between the model and human experts) increased from 3.2% to 12.7%, and the F1-Score decreased by 2.83%. The Cluster Separation Index (CSI) is used to monitor excessive reward optimization, and an early stopping mechanism is triggered when CSI > 0.8.

3) Dynamic Adjustment of KL Penalty Term

The reference model uses a frozen SFT-BERT. The KL loss weight decreases linearly from 0.2 to 0.05 with the number of iterations to balance model exploration and stability.

F. MODEL TRAINING AND INFERENCE PROCESS

(1) Pre-training Phase: Both the Actor and Reward Model use BERT fine-tuned with power domain corpus as the initial weight;
(2) Reinforcement Learning Phase: The PPO algorithm is used to update the Actor. The advantage function window size is set to 5, the learning rate is set to 2×10^{-5}, and 10 training are performed after each round of feedback;
(3) Transition from Reinforcement Learning to Inference Phase: After completing reinforcement learning, the model parameters are frozen to ensure inference stability. Real-time logs undergo the same preprocessing steps as the training data (log cleaning, terminology standardization, and feature embedding).
(4) Inference Phase: After preprocessing, real-time logs are matched with the template library and inherit labels. Unmatched samples trigger online knowledge retrieval and dynamic prompt generation, and finally output hazard types and confidence levels.

IV. EXPERIMENTAL VERIFICATION AND RESULT ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND DATASET

1) Experimental Environment

Hardware: Intel Xeon Gold 6338 processor, 4×NVIDIA A100 GPUs, 1 TB memory.

Software: PyTorch 2.1, Hugging Face Transformers, DeepPavlov framework.

2) Dataset Setup

The 8,632 labeled samples described above are divided into a training set (6,042 samples), a validation set (1,726 samples), and a test set (864 samples) in a 7:2:1 ratio. For small-sample experiments, 10%, 20%, and 50% of the training set samples are used for comparison.

B. BASELINE MODELS AND EVALUATION METRICS

1) Baseline Models

(1) BERT-BiLSTM-CRF: A mainstream model for power text entity recognition;
(2) SFT-BERT: A domain-adapted model only fine-tuned via supervised learning;
(3) Standard RLHF: A basic framework without introducing APO and InfoRM;
(4) ERNIE-3.0-Base: A knowledge-enhanced pre-trained model developed by Baidu, fine-tuned with power domain corpus;
(5) ChatGLM-6B (SFT): A power hazard identification model fine-tuned based on a dialogue model.
(6) GPT-3.5-Turbo (SFT): A GPT-style model fine-tuned with the same power dispatching log corpus to isolate the impact of the RLHF component.

2) Evaluation Metrics

(1) Precision: Measures the reliability of identification results, with the formula as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{1}$$

True Positive (TP): The number of samples that are classifies as hazardous by the model and are actually hazardous.

False Positive (FP): The number of samples that are classifies as hazardous by the model but are actually non-hazardous.

(2) Recall: Measures the completeness of hazard identification, with the formula as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

False Negative (FN): The number of samples that the model classifies as non-hazardous but are actually hazardous.

(3) F1-Score: Measures the balance between precision and recall, with the formula as follows:

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\% \tag{3}$$

(4) Generalization Error (GE): Measures the adaptability in small-sample scenarios, with the formula as follows:

$$GE = \left(1 - \frac{F1_{\text{small subset}}}{F1_{\text{full set}}}\right) \times 100\% \tag{4}$$

$F1_{\text{small-sample}}$: The F1-Score on the test set when the model is trained using a portion of the training set samples (e.g., 10%, 20%, 50%). $F1_{\text{full-sample}}$: The F1-Score on the test set when the model is trained using all samples (100%) of the training set.

(5) Hazard Level Prediction Accuracy (HLPA): Measures the precision of the model in the three-level classification of major/general/non-hazard, with the formula as follows:

$$HLPA = \frac{\sum_{i=1}^{3} TP_i}{\sum_{i=1}^{3}(TP_i + FP_i + FN_i)} \times 100\% \quad (5)$$

TP_i represents the number of true positives for the i-th level hazard; FP_i represents the number of false positives for the i-th level hazard; FN_i represents the number of false negatives for the i-th level hazard.

(6) Entity Extraction F1-Score (Entity-F1): Calculates the F1-Score for entity extraction, targeting key entities in logs such as equipment names, parameter values, and operational behaviors.

C. EXPERIMENTAL RESULTS AND EXTENDED ANALYSIS

1) Enhanced Comparison of Overall Performance

As shown in Table 2, the model proposed in this paper achieves the best performance across all metrics: Its F1-Score is 1.6% higher than that of the second-best model (Standard RLHF) and 3.21% higher than that of ERNIE-3.0-Base. The HLPA reaches 95.3%, solving the key issue of major hazards being misjudged as general hazards in traditional models. The Entity-F1 reaches 94.8%, proving that domain knowledge embedding effectively improves the precision of entity extraction.

TABLE 2. Multi-Metric Performance Comparison of Multiple Models (Test Set)

Model	Precision (%)	Recall (%)	F1-Score (%)	HLPA (%)	Entity-F1 (%)	Inference Time (ms/sample)
BERT-BiLSTM-CRF	92.54	92.05	92.30	89.7	88.2	87
ERNIE-3.0-Base	93.42	92.81	93.11	91.2	90.5	63
ChatGLM-6B (SFT)	94.15	93.27	93.71	92.5	91.3	124
SFT-BERT	93.17	91.82	92.49	90.1	89.3	42
Standard RLHF	95.36	94.01	94.68	93.8	92.1	68
GPT-3.5-Turbo (SFT)	94.89	93.65	94.27	93.1	91.8	56
Proposed Model	96.82	95.74	96.28	95.3	94.8	75

Note: The inference time of 75ms/sample for the proposed model was verified in a real-world provincial-level power dispatching center environment. The center processes approximately 5,000 real-time logs per minute, and the total inference time per minute is 5,000 × 75 ms = 375 s, which is far below the 600s (10-minute) response threshold required by the center. This confirms that the model meets the real-time monitoring requirements for massive stream data.

2) In-depth Analysis of Multi-Round Feedback Effects

As shown in Table 3, all metrics improved significantly after three rounds of feedback, with diminishing marginal benefits:

From Round 1 to Round 2: The F1-Score increased by 3.34%, mainly due to the APO strategy screening high-value samples and solving the problem of ambiguous semantic recognition;

From Round 2 to Round 3: The F1-Score increased by 2.13%, focusing on optimizing confusion between similar terms. For example, the recognition precision of vertical encryption and horizontal isolation increased from 78% to 96%;

After the 4th round of feedback, the metrics basically stabilized, with the F1-Score increasing by only 0.02%. This demonstrates that three rounds of feedback represent the optimal number of iterations, balancing performance and labor costs.

TABLE 3. Variation Trends of Metrics Across Multi-Round Feedback

Feedback Round	Precision (%)	Recall (%)	F1-Score (%)	HLPA (%)	Entity-F1 (%)	Labor Time (Hours)
Initial Model	88.62	87.35	87.98	86.4	85.7	0
Round 1	91.35	90.27	90.81	89.2	87.9	4.2
Round 2	94.72	93.58	94.15	92.9	91.5	7.8
Round 3	96.82	95.74	96.28	95.3	94.8	10.5
Round 4	96.79	95.81	96.30	95.4	94.9	14.1

3) In-Depth Analysis of Error Cases

Error cases are divided into 4 categories, with supplementary specific data and optimization directions shown in Table 4.

TABLE 4. Classification and Analysis of Error Cases

Error Type	Proportion (%)	Case Example	Error Cause	Optimization Direction
Insufficient New Hazard Samples	38	Phase A communication interruption at the new energy station, delayed data upload	The training set contains only 42 new hazard samples, resulting in insufficient feature learning	Build an online learning module to supplement new hazard samples in real time
Confusion Between Similar Terms	27	AC control and protection device failure was misjudged as DC control and protection failure	The semantic similarity of terms reaches 0.89, making it impossible for the model to distinguish	Optimize the knowledge graph and add term attribute features (e.g., DC/AC labels)
Context-Dependent Errors	22	No supervision implemented during the operation of 110 kV Xinling Line was misjudged as operational hazard	Failed to combine the process logic of supervision required before operation and ignored contextual correlation	Introduce power dispatching process knowledge and build a contextual reasoning module
Ambiguous Parameter Thresholds	13	Acetylene content of 0.3 μL/L was misjudged as transformer gas anomaly	Did not match the 0.5 μL/L threshold specified in Guidelines for Analysis and Judgment of Dissolved Gases in Transformer Oil	Build a parameter threshold knowledge base to realize automatic matching between text parameters and standard thresholds

V. CONCLUSIONS

The RLHF multi-round human feedback model proposed in this paper effectively improves the precision of hazard identification in power dispatching logs through domain knowledge embedding, APO sample selection, and InfoRM reward optimization, thereby enhancing system stability against operational risks posed by energy-intensive clients such as the industry. The model has the following advantages:

(1) Performance Advantage: The proposed model outperforms mainstream baseline models in core metrics such as precision, F1-Score, and HLPA;
(2) Efficiency Advantage: The multi-round feedback mechanism increases the efficiency of manual labeling by 3.5 times. Three rounds of feedback are the optimal number of iterations, balancing performance and cost;
(3) Module Effectiveness: The APO strategy, InfoRM mechanism, and domain knowledge embedding all significantly improve performance;
(4) Practical Adaptability: The model has good stability (F1 fluctuation of 0.7%), and its inference time meets the requirement of second-level response, making it adaptable to actual power dispatching scenarios.

AUTHOR CONTRIBUTIONS

Conceptualization – Siwu Yu, Yumin He, Guobang Ban, Jintong Ma, Guanghui Xi, Lingwen Meng, Shasha Luo and Siqi Guo; methodology – Siwu Yu, Yumin He, Guobang Ban, Jintong Ma, Guanghui Xi, Shasha Luo and Siqi Guo; investigation – Siwu Yu, Yumin He, Guobang Ban, Jintong Ma, Guanghui Xi, Lingwen Meng, Shasha Luo and Siqi Guo; writing-original draft preparation – Siwu Yu, Yumin He, Guobang Ban, Jintong Ma, Guanghui Xi, Lingwen Meng, Shasha Luo and Siqi Guo. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This work is financially supported by the Science and technology project of China Southern Power Grid Co., Ltd. (No.GZKJXM20232503).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. L. Tu, "Analysis and Processing of Power System Log Data Based on Spark," Southeast University, 2017.

[2] Y. B. Shu, Y. Tang, and H. D. Sun, "Research on safety and stability standards for power systems," Proceedings of the CSEE, vol. 33, no. 25, p. 8, 2013.

[3] P. Zhang, H. F. Yang, and Y. B. Xu, "Power big data and its applications in power grid companies," Proceedings of the CSEE, vol. 34, no. S1, pp. 85–92, 2014, doi: 10.13334/j.0258-8013.pcsee.2014.S.012.

[4] J. Zhang, "Current status and challenges of big data processing technology for smart grids," Power System Technology, vol. 37, no. 4, p. 9, 2013, doi: 10.3969/j.issn.1009-8119.2015.10.099.

[5] B. H. Lü, K. Sun, Z. T. Fan, Y. Guo, and F. Yang, "A refined management platform for power dispatching standard records based on abnormal operation accident data," Electrical Application, vol. 43, no. 10, pp. 47–53, 2024.

[6] W. J. Luo, F. Qian, H. Y. Zhu, J. X. He, Y. K. Huang, Y. H. He, et al., "A method and device for automatic filling of dispatching logs," 2022.

[7] Z. Ni, Q. M. Li, and Y. J. Guo, "An integrated prediction algorithm for anomaly monitoring in power big data log analysis platforms," Journal of Nanjing University of Science and Technology, vol. 41, no. 5, p. 12, 2017, doi: 10.14177/j.cnki.32-1397n.2017.41.05.016.

[8] Y. Lu and J. M. Sun, "Research on log analysis technology based on big data and its application in enterprise information security," China High-Tech, no. 18, pp. 7–9, 2022, doi: 10.13535/j.cnki.10-1507/n.2022.18.01.

[9] C. Hu, C. L. Liu, and S. Wu, "Analysis of early warning technology based on big data logs," Electrical Technology, no. 6, p. 4, 2017, doi: 10.3969/j.issn.1673-3800.2017.06.021.

[10] Y. J. Yan, G. H. Sheng, Y. F. Chen, X. C. Jiang, Z. H. Guo, and X. M. Du, "Anomaly detection method for power transmission and transformation equipment status data based on big data analysis," Proceedings of the CSEE, vol. 35, no. 1, 2015, doi: 10.13334/j.0258-8013.pcsee.2015.01.007.

[11] W. Xing, J. Y. Lang, Y. Qu, W. J. Yao, X. Y. Zhang, S. Y. Zhang, et al., "A big data transmission system based on power user behavior logs," 2023.

[12] X. Q. Yu and L. H. Qi, "Anomaly detection of power big data based on stream data clustering algorithm," Power Information and Communication Technology, vol. 18, no. 3, pp. 8–14, 2020, doi: 10.16543/j.2095-641x.electric.power.ict.2020.03.00.

[13] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, et al., "Training language models to follow instructions with human feedback," arXiv e-prints, 2022.

[14] W. B. Yang and X. Li, "Research on a collaborative proofreading method for large language models based on BERT and RLHF," Modern Information Technology, no. 11, 2025, doi: 10.19850/j.cnki.2096-4706.2025.11.008.

[15] J. Zhao, F. Wen, and Y. Xue, "Architecture, implementation technologies and challenges of power CPS," Automation of Electric Power Systems, vol. 34, no. 16, pp. 1–7, 2010.

[16] C. R. Green, L. Wang, and M. Alam, "Applications and Trends of High Performance Computing for Electric Power Systems: Focusing on Smart Grid," IEEE Transactions on Smart Grid, 2013, doi: 10.1109/TSG.2012.2225646.

[17] X. M. Ye, J. H. Zhao, and F. S. Wen, "Quantitative assessment of vulnerability in power information systems based on adjacency matrix," Automation of Electric Power Systems, vol. 37, no. 22, pp. 41–46, 2013, doi: 10.7500/AEPS20130324004.

[18] B. Babcock, M. Datar, and R. Motwani, "Load Shedding for Aggregation Queries over Data Streams," in Proceedings of the 20th International Conference on Data Engineering, 2004, doi: 10.1109/ICDE.2004.1320010.

[19] M. Shi, "Research on Application of Association Rules Based on Clustering Partition in Web Log Mining," Wuhan University of Technology, 2014, doi: 10.7666/d.D617025.

[20] W. Fu, "Design and Implementation of a Web Log Analysis Platform Based on Hadoop," Beijing University of Posts and Telecommunications, 2015.

[21] Y. Liang, Y. Zhang, H. Xiong, and R. Sahoo, "Failure Prediction in IBM BlueGene/L Event Logs," in ICDM, 2007, doi: 10.1109/ICDM.2007.46.

[22] P. Gujrati, Y. Li, Z. Lan, R. Thakur, and J. White, "A Meta-Learning Failure Predictor for Blue Gene/L Systems," in International Conference on Parallel Processing, 2007, doi: 10.1109/ICPP.2007.9.

[23] Z. Lan, J. Gu, Z. Zheng, R. Thakur, and S. Coghlan, "A study of dynamic meta-learning for failure prediction in large-scale systems," Journal of Parallel & Distributed Computing, vol. 70, no. 6, pp. 630–643, 2010, doi: 10.1016/j.jpdc.2010.03.003.

[24] I. Fronza, A. Sillitti, G. Succi, M. Terho, and J. Vlasenko, "Failure prediction based on log files using Random Indexing and Support Vector Machines," Journal of Systems and Software, vol. 86, no. 1, pp. 2–11, 2013, doi: 10.1016/j.jss.2012.06.025.

[25] N. Nakka, A. Agrawal, and A. Choudhary, "Predicting Node Failure in High Performance Computing Systems from Failure and Usage Logs," in 2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and PhD Forum, 2011, pp. 1557–1566, doi: 10.1109/IPDPS.2011.310.

[26] Y. Bang, S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, et al., "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity," 2023.