

## Research Article

# An Association Rule Mining-Based Method for Revealing the Impact of Operational Sequence on Nuclear Power Plants Operating

Yuxuan He <sup>1</sup>, Jian Song <sup>1</sup>, Shaoke Shi,<sup>1</sup> Haibo Lian,<sup>1</sup> Jiangyang He,<sup>1</sup> Ren Yu <sup>2</sup>, Tete Liu,<sup>1</sup> Bin Sun,<sup>1</sup> Jiangtao Yuan,<sup>1</sup> and Yingbin Hu<sup>1</sup>

<sup>1</sup>Navy Submarine Academy, Qingdao, Shandong 266000, China

<sup>2</sup>Naval University of Engineering, Wuhan, Hubei 430033, China

Correspondence should be addressed to Jian Song; [jian\\_s@hotmail.com](mailto:jian_s@hotmail.com)

Received 6 June 2023; Revised 22 February 2024; Accepted 18 March 2024; Published 25 March 2024

Academic Editor: Han Zhang

Copyright © 2024 Yuxuan He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The operations of the operators are important for nuclear safety, but conventional operating experience feedback and common data-driven methods make it difficult to explicitly find valuable information hidden in these operational sequences that can help the operator to provide advice at the operational level. During the nuclear power plant (NPP) operation, a large amount of historical operating data is accumulated, which records the operational sequences of the operators and the state parameters of equipment. Therefore, this paper proposes the use of association rule techniques to mine the NPP operating data to discover the operational characteristics of operators and reveal their possible impact on the NPP operation. This work helps to improve the operational performance of operators and prevent human-factor events. To this end, the concept of state switching values for describing the operating states of NPPs is proposed to enable the proposed method to be adapted to different practical application scenarios. A sequence segmentation method is proposed to be able to transform historical NPP operating data into a sequence data set for association rule mining. Furthermore, an ensemble algorithm based on sequence pattern mining and sequence rule mining and its postprocessing method are designed. The empirical study was carried out using 20 batches of historical operating data of the cold start-up. A total of 164 original association rules are generated using the proposed method and were analyzed by experts. The recommendations were made for 4 different cases that would improve the operational performance of the operators.

## 1. Introduction

With the continuous increase in the number of operating experience feedback (OEF) from nuclear power plants (NPPs), analysis of human factors events (HFEs) has become an important part of OEF and is important for improving personnel performance and preventing human errors [1, 2]. Many studies have been conducted on HFEs in recent years [3–6]. It is evident that human reliability analysis (HRA) tends to focus more on the impact of objective conditions on the operator and lacks analysis of the operational level of the operator [7, 8]. Moreover, the main source of HRA data for the study is full-scope simulators; operational level problems are more likely to occur under real operating conditions [9].

In HFEs, the operations of the operator have a direct impact on nuclear safety and account for a large proportion of human errors [10]. Faced with an emergency situation, operators may make incorrect judgements about the state of the equipment and perform corresponding operations accordingly [11]. More often, due to variations in cognitive ability, experience level, and behaviour among different operators at nuclear power plants, operational sequences may display differing characteristics, such as regular performance of certain operations and habitual disregard for procedural steps [12], even if the initial state of the NPP is the same. Revealing hidden characteristics in these operational sequences can help prevent human error and improve the safety of NPPs.

Currently, there are two approaches for analysing the operational sequences of NPP operators. One approach involves obtaining OEF subsequent to anomalies or accidents. However, the frequency of such cases is very low. For operations that do not generate anomalies or accidents, it becomes arduous to accurately determine the consequences for NPPs. In this context, if operators ignore characteristics that are hidden in operational sequences, the probability of HFEs is likely to increase. Another approach is the presently popular data-driven techniques such as neural networks, support vector machines, clustering, and others [13]. These techniques can extract operational characteristics of operators from data. Data-driven techniques have extensively been utilized in condition monitoring, fault diagnosis, detection of environmental radiation, and other related domains of NPPs [14]. The issue with this approach is that the model's extracted features from data are often implicit and challenging to interpret, thus insufficient in aiding operators to enhance their operational performance for a particular case. Furthermore, the model's generalization performance is limited due to the nature of the data, resulting in a situation where implementing a data-driven model in complex real-life scenarios is unfavourable for obtaining dependable operation recommendations.

Association rule mining is one of the main techniques of data mining, which can find frequent patterns, associations, and correlations among variables or items in a database [15]. Unlike the data-driven approach mentioned above, the method of obtaining frequent items in the data is explicit and does not involve model learning. This method can provide results in the form of rules that are intuitive and easy to understand. In addition, it exhibits strong applicability to various data sets and does not require excessive parameter tuning. These characteristics of association rules make them able to be directly used to mine the operational patterns of operators from historical data, and effective conclusions and suggestions can be obtained through expert analysis. There has been some research into the use of association rule mining techniques for HFEs in NPPs. Jiang et al. used an association rule-based approach to assess the support and confidence level of HFEs, using the example of a steam generator tube rupture accident, to assist in reducing HFEs [16]. Zou et al. used association rule techniques to identify associations and causality among HFEs [1]. These studies have inspired us to consider using this technology to analyze the association between the operational sequence and some operating phenomena to help the operator improve operational performance and prevent human errors. In addition, with the increasing digitization of NPPs, a large amount of historical operating data is now available [17]. These data contain information on the condition of the equipment and the operations of the operators, and already provide the data basis for carrying out related research.

This paper proposes a method based on association rule mining to discover the operational sequence characteristics of operators through historical operating data of NPPs and reveal the impact of these operational sequences on NPPs. The proposed method conducts an association analysis on whether these operational sequences are conducive to the normal operation of NPPs.

## 2. Methodology

The outline of the proposed method is shown in Figure 1. It consists of three modules, which are data preprocessing, association rule mining, and rule postprocessing. Data preprocessing aims to generate state switching values that represent operation events and to segment the raw data into sequence datasets accordingly. Association rule mining is the central part, the ultimate aim of which is to discover original rules between operational sequences and operation events represented by the state switching values. The purpose of rule postprocessing is to obtain rules and accurate confidence for expert analysis and to reveal the impact of operational sequence on NPPs operating according to these rules and metrics. In Section 2.1, state switching values are proposed to adapt the association rule technique to varying task types. In addition, a sequence segmentation method is presented that accounts for the specific features of data obtained from the NPPs operating data. In Section 2.2, the technical principles of association rules are introduced, and our proposed approach is shown. In Section 2.3, we outline the process of revising association rule mining results to obtain a form suitable for expert analysis.

*2.1. Data Preprocessing.* An NPP has many different types of systems and equipment, and its data acquisition devices acquire and record thousands of major parameters [18]. These data variables can be divided into analog quantities and switching values. Analog quantities are some physical quantities that vary continuously over time, usually represented by numerical values with units. The state of the NPP system, such as pressure and flow, can be monitored by observing the changes in the analog quantities during operation. The switching value is a physical quantity with only two states, usually represented by 0 and 1. The switching value can be further divided into a control switching value and alarm switching value. The control switching value reflects the intervention actions of the operator, e.g., the stop of a pump can be represented by  $1 \rightarrow 0$ , and the action of a control rod can be represented by  $0 \rightarrow 1$ . The alarm switching value is a variable that reflects whether the key operating parameter exceeds the safety limit. The original dataset for the method proposed in this paper is the time series data containing the above variable types, which is transformed into a segmented sequence dataset containing control switching values and state switching values after data preprocessing. For the convenience of understanding, Figure 2 shows how the proposed method changes the format of the data during process.

*2.1.1. State Switching Value Generation.* Since the alarm switching value describes the abnormal or accident state of the NPP in terms of a definite safety limit, the information in the period before and after the alarm switching value is triggered will be lost. Moreover, the scenarios for triggering the alarm switching value in actual operation are very limited. These factors are not conducive to data analysis and mining. Based on the prior knowledge of experts, this paper

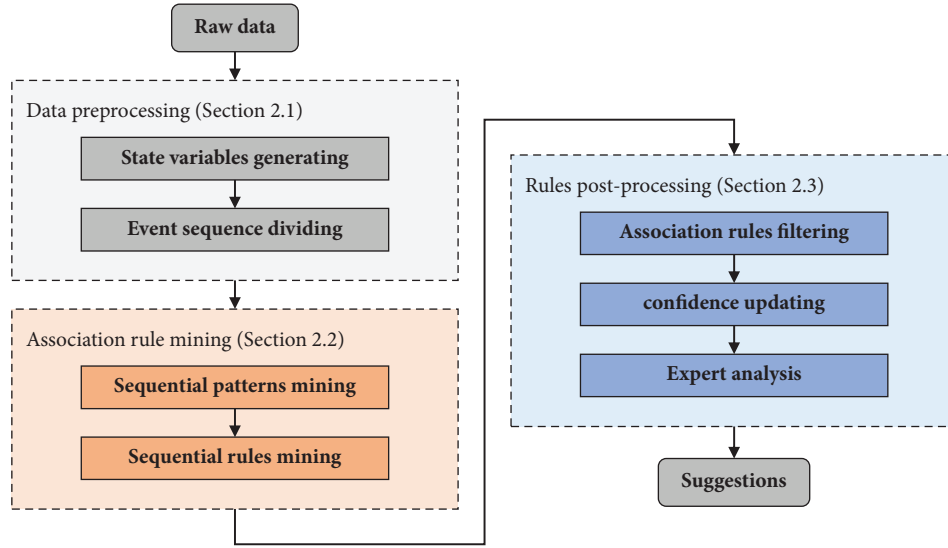


FIGURE 1: Outline of the proposed method.

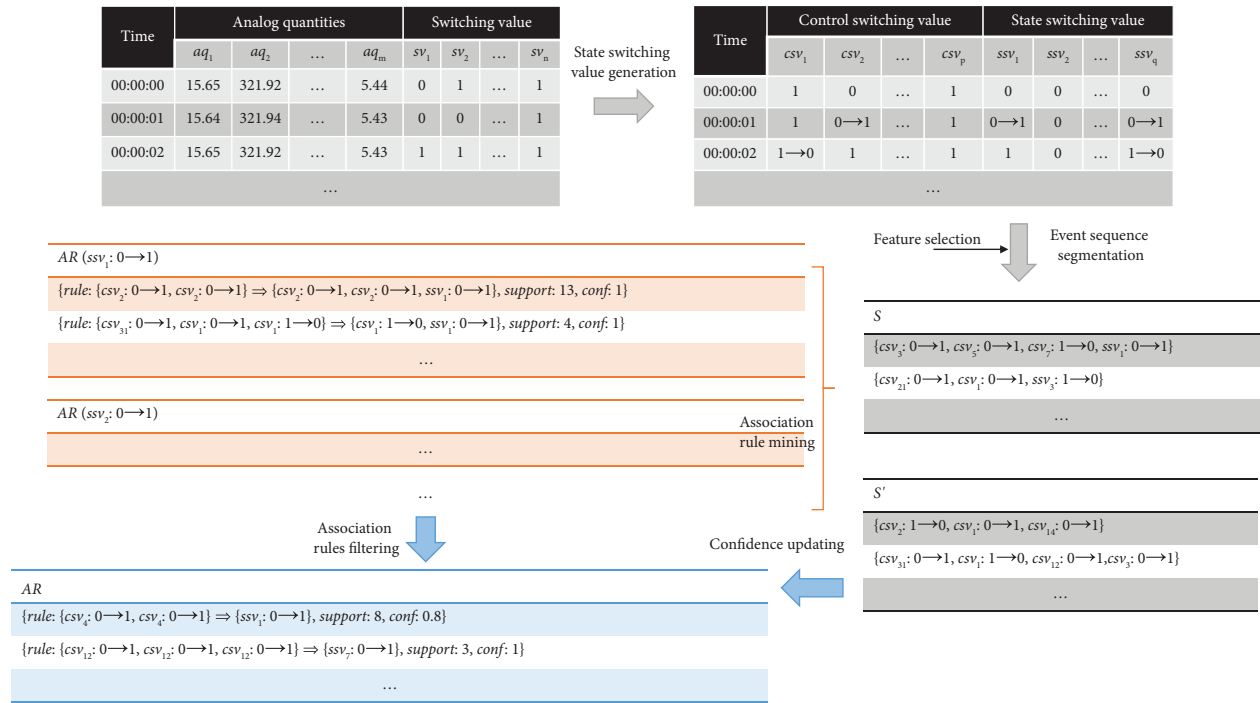


FIGURE 2: The change of data format during the proposed method.

proposes the concept of state switching value, which is used to find the sample points that may tend to or are in an abnormal or accident state. The state switching value is a variable that describes whether the value of continuous variables such as temperature and pressure is in a state conducive to the normal operating of the NPP from the perspective of data distribution and is represented by 0 and 1. Where 0 represents the normal operating state, 1 represents the deviation from the normal operating state, and 0  $\rightarrow$  1 and 1  $\rightarrow$  0 indicate the state change. The state switching values are defined as follows.

*Threshold definition method.* Arrange the values of the continuous variable  $x$  in ascending or descending order, and then, according to the physical meaning of the variable, mark the first  $n\%$  of the variable  $x$  as state 1, and the rest as state 0.

*Differential definition method.* Arrange the continuous variables  $x$  in chronological order, set the value of the variable  $x$  at the current moment  $t_1$  to  $q_1$ , and the value of the variable  $x$  at the previous moment  $t_0$  to be  $q_0$ , and perform the following calculations for each moment:

$$x_d = \frac{q_1 - q_0}{t_1 - t_0}. \quad (1)$$

The resulting  $x_d$  is then arranged in ascending or descending order, marking the first  $n\%$  of moments of the variable  $x_d$  as state 1 and the rest as state 0, depending on the physical meaning of the variable.

*Moving average definition method.* For continuous variables  $x$  in time order, let the value of variable  $x$  at the current moment be  $q_m$  and the values of variable  $x$  at the previous  $m - 1$  moments be  $q_{m-1}, \dots, q_1$ , respectively. For each moment the following calculation is performed:

$$x_m = \frac{q_1 + q_2 + \dots + q_m}{t_m - t_1}. \quad (2)$$

Then sort  $x_m$  in ascending or descending order, according to the physical meaning of the variable, and mark the first  $n\%$  of variable  $x_m$  as state 1, and the rest as state 0.

The threshold definition method finds the states within the statistics that are closer or further away from the safety limit by applying constraints between the upper and lower limits of the alarm switching threshold. The differential definition method is used to observe states where the variables change instantaneously, either more rapidly or more slowly. The moving average definition method is suitable for analysing the parameters which are prone to drastic fluctuations and identifying the variables with strong nonlinear changes over some time. In practice, irrelevant and redundant variables can be removed in advance using a filter feature selection algorithm based on a defined state switching value to improve mining efficiency and reduce postprocessing difficulties [19].

*2.1.2. Event Sequence Segmentation.* Before mining association rules, the event sequence should be segmented into sequence datasets. The commonly used static time series segmentation methods are piecewise aggregate approximation (PAA) [20], discrete Fourier transform (DFT) [21, 22], symbolic aggregate approximate (SAX) [23], and special points-based method [24, 25], etc. However, the process of segmenting the historical operating data of NPPs based on state switching values has the following difficulties.

- (i) Segmentation by time intervals only would result in different events appearing in the same sequence at the same time and affecting each other. Moreover, if different events are selected as research objects, different segmentation results will be obtained.
- (ii) If the time interval is set too long, it can lead to irrelevant events appearing in the same sequence and may produce some pseudo-rules with low support and confidence. In addition, setting the time interval too short can lead to some events that would otherwise be associated with being divided into different sequences, resulting in low confidence in the rule.

- (iii) There may be interference sequences in the event sequence, such as  $0 \rightarrow 1$  and  $1 \rightarrow 0$  that change frequently in the state switch value in a short time.
- (iv) If the segmented dataset has extra-long sequences, the efficiency of the sequence segmentation algorithm can be seriously affected.
- (v) If the whole sequence dataset is used as input to the association rule mining algorithm after segmentation, it will be too computationally intensive to complete the association rule mining algorithm.

Aiming at the above difficulties, a sequence segmentation method is designed in this paper to target the event sequence. The motivation for proposing this approach is that the issues mentioned above are of an engineering nature rather than an academic nature. Existing methods struggle with the need to efficiently generate segmentation sequences with different time window sizes and event sequences of different lengths from raw time series data. Therefore, the idea of event sequence segmentation is to consider only the moments where there is a change in the state switching values and to output them in two different sets according to whether they contain state switching values or not. Only the set containing state switching values is used for association rule mining.

Figure 3 shows the sequence segmentation process using only a single state switching value change as an example. The input data set  $D$  is the time series data, including the timestamp, the control switching values at each time, and the change of the state switching values ( $0 \rightarrow 1$  or  $1 \rightarrow 0$ ). The process scans the database only once and records the moments and operational sequences where switching values change. The two parameters  $\lambda$  and  $\mu$  are used to limit the sequence length and the event interval, facilitating expert analysis during rule postprocessing. With this method, two sets of sequences  $S$  and  $S'$  are generated for each state switching value, where  $S$  is used as input to the association rule mining algorithm and  $S'$  is used to update the confidence in rule postprocessing. When segmenting the input with multiple state switching value changes, only parallel processing is required.

*2.2. Association Rule Mining.* As it is the sequence data with a timestamp that reflects the operator's operational sequence, the association rule mining module uses sequential pattern mining and sequential rule mining techniques to process the segmented sequence dataset.

Sequential pattern mining was proposed by Agrawal et al. to mine frequently occurring ordered events or subsequences as patterns [26]. The problem of sequential pattern mining can be briefly stated as follows [27].

Let  $I = \{i_1, i_2, \dots, i_k\}$  be the set of  $k$  items. The sequence  $\alpha = \langle a_1, a_2, \dots, a_m \rangle$  ( $a_i \in I$ ) is an ordered list in which each set of items represents events that occur at the same timestamp. For a sequence  $\beta = \langle b_1, b_2, \dots, b_n \rangle$ , we call  $\alpha$  a subsequence of  $\beta$  if and only if there exists  $i_1, i_2, \dots, i_k$  such

---

**Input:**  $D = \{time, f_1, \dots, f_n, state\ change\}$  // time sequence data  
 $\lambda$  // maximum time interval  
 $\mu$  // maximum sequence length

**Output:**  $S = \{operational\ sequence, state\ change\}$   
 $S' = \{operational\ sequence\}$

---

```

1: for all times  $D_i \in D$ , with  $D_i \neq Null$  do
2:    $sequence = getSequence(start\ time, current\ time)$ 
3:    $time\ interval = current\ time - start\ time$ 
4:   if ( $time\ interval > \lambda$  and  $sequence$ ) then
5:     append  $sequence$  to  $S'$  and clear  $sequence$ 
6:   if ( $state\ change \neq Null$  and  $sequence$ ) then
7:     append  $state\ change$  to  $sequence$ 
8:     if ( $len(sequence) < \mu$ ) then
9:       append  $sequence$  to  $S$ 
10:    else then
11:      append  $getLasterSequece(sequence, \mu)$  to  $S$ 
12:      append  $getFormerSequence(sequence, \mu)$  to  $S'$ 
13:      clear  $sequence$ 
14:     if ( $sequence == Null$ ) then update  $start\ time$ 
15:   end for

```

---

FIGURE 3: Sequence segmentation method.

that  $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$  and  $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_m \subseteq b_{i_m}$ , denoted as  $\alpha \sqsubseteq \beta$ . For a given sequence data set  $S = \{s_1, s_2, \dots, s_n\}$ , the support of a sequence  $\alpha$  is the number of  $\alpha$  contained in  $S$ . If the sequence  $\alpha$  satisfies the minimum support threshold,  $\alpha$  is a sequential pattern.

The key to association rule mining is to find all sequential patterns efficiently. In this paper, the idea of prefix-projected pattern growth proposed by Pei et al. is used to traverse the search space to enumerate all frequent sequences [28]. The main ideas are as follows: (i) The first scan of the database yields a set of sequential patterns of length 1. (ii) Each sequential pattern is regarded as a prefix, and the complete set of sequential patterns can be divided into different subsets depending on the prefix. (iii) To mine a subset of the sequential patterns, the corresponding projection database is constructed and recursively mined. Given the minimum support  $\gamma$ , the above process can be performed to obtain frequent terms, as shown in Figure 4 (lines 1–13).

Once the frequent sequences are known, they can be used to obtain rules describing the relationships between different sequence items [29]. In this paper, we use the association rule representation in [30], i.e.,  $A \implies B$ . The confidence of  $A$  is expressed as  $fr(A \cup B)/fr(A)$  and  $fr(\cdot)$  represents the frequency of sequence occurrence. Given the minimum confidence, we use the algorithm shown in Figure 4 (lines 14 to 18) to generate rules that satisfy the conditions. The rules have the form  $A \implies B$  as described above, where  $A$  is a subsequence of  $B$ .

Figure 4 shows the ensemble algorithm for association rule mining. The data input of the algorithm is a set of segmented sequences according to the state switching values, and the set of segmented sequences with different state switching values can be combined as the whole input. Lines

1–13 represents the process that searches for sequential patterns, and the result is a set  $F$  composed of sequential patterns. Lines 14–18 is the process by which generates sequence rules based on the set  $F$ . The result is a set AR consisting of sequence rules, support, and confidence shown in Figure 2 (orange table). Due to the limitations of the output data format, the minimum confidence  $\eta$  of the algorithm needs to be set to 1 to find the sequence of operations associated with the state switching values, and the true confidence is updated in the rule postprocessing.

**2.3. Rule Postprocessing.** After association rule mining, each event corresponding to the change of state switching value ( $0 \rightarrow 1$  or  $1 \rightarrow 0$ ) generates a rule set to be updated. Association rules for expert analysis are finally obtained through association rule filtering and confidence updating. For expert analysis, each action in the operational sequence carries a timestamp, as shown in Figure 5.

**2.3.1. Association Rules Filtering.** For rule  $A \implies B$ , we call rule  $A$  an antecedent sequence and rule  $B$  a consequence sequence. The final purpose of association rule filtering is to make the antecedent sequence the operational sequence of the operator, and the consequence sequence the sequence that contains the sequence of operations and the event of a change in a state switching value after the sequence of operations has occurred. However, many of the rules generated by association rule mining may not satisfy the above purposes. Therefore, in this phase of association rule filtering, it is considered to filter the rules for each state transition set using the data format requirements. Filter the judged rule when the following occurs.

---

**Input:**  $S = \{\text{operational sequence, state change}\}$  // sequence data  
 $\gamma$  // minimum support  
 $\eta$  // minimum confidence

**Output:**  $AR = \{\text{rule, support, conf}\}$

---

```

1: for all actions  $A_k \in S$ , with  $\text{len}(A_k) == 1$  do
2:   if ( $\text{count}(A_k) < \gamma$ ) then delete  $A_k$ 
3:   else then  $P_k = \text{getNextProjection}(A_k)$ 
4: end for
5: for all remained  $A_k$  do
6:   if ( $P_k == \text{Null}$  or  $\text{countEachItem}(P_k) < \gamma$ ) then continue
7:   else then
8:      $P_k = \text{removeLowItem}(P_k)$ 
9:      $A_k = \text{merge}(A_k, P_k)$ 
10:     $P_k = \text{getNextProjection}(P_k)$ 
11:    if ( $P_k == \text{Null}$ ) then append  $A_k$  to  $F$ 
12:    else then continue
13:  end for
14: for all frequent sequences  $A_k \in F$  do
15:   for all subsequences  $B_k < A_k$  do
16:     $\text{confidence} = \text{fr}(B_k) / \text{fr}(A_k)$ 
17:    if ( $\text{confidence} \geq \eta$ ) then output the rule  $B_k \Rightarrow A_k$ , support, and confidence
18:  end for

```

---

FIGURE 4: Ensemble algorithm for association rule mining.

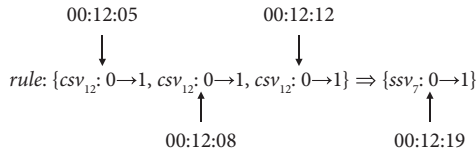


FIGURE 5: Association rule with timestamp.

- (1) Length (antecedent sequence) + 1  $\neq$  length (consequence sequence)
- (2) Pop (consequence sequence)  $\neq$  antecedent sequence
- (3) Length (consequence sequence[end])  $\neq$  1
- (4) Consequence sequence[end]  $\neq$  current state switching value
- (5)  $\begin{cases} \text{Support}(\text{Rule\_judged}) = \text{support}(\text{Rule}), \\ \text{Confidence}(\text{Rule\_judged}) = \text{confidence}(\text{Rule}). \end{cases}$  and  $\text{Rule\_judged} = \text{subsequence}(\text{Rule})$

**2.3.2. Confidence Updating.** Support and confidence are two important metrics of association rules. If support is too low, it means the rule is highly contingent. When the confidence is too low, the antecedent and consequence of the rule are associated to a low degree. The confidence updating aims to obtain the true confidence of the filtered rule through the sequence set  $S'$  which does not contain the changes of state switching value generated in the process of event sequence segmentation. For the convenience of subsequent expert analysis, the part of the consequence sequence that is

unrelated to the state switching value change events is removed, i.e., rules are represented by the *operational sequence*  $\Rightarrow$  *state switching value changes*.

The process of confidence updating is shown in Figure 6. By executing this process on the sequence ( $S'$ ) without state switching value change and the corresponding association rule (AR) generated by each state switching value change, the association rule with updated confidence can be obtained. In this process, the parameter  $\varepsilon$  is used as the minimum confidence of the truth, and finally, the association rules that meet the requirements are obtained.

### 3. Empirical Study

To verify the effectiveness of the method proposed in this paper, the historical data of the cold start-up were used to carry out empirical research, analyze the operational sequence of operators, and put forward guidance opinions that can help improve the performance of operators and avoid HFEs [31]. The cold start-up of the NPP is the process of moving a nuclear reactor from a cold state to stable power. It can be roughly divided into 3 stages [32]: (1) system preparation, (2) subcritical state to critical state, and (3) heat-up. This process involves the extensive operation of pumps, valves, electrical heaters, and control rods and requires consideration of the appropriate coordination of the operating states of various subsystems [31–33]. At present, the start-up control of reactors is mainly realized manually by operators, which takes a long time and places a heavy burden on operators, which is prone to HFEs.

---

```

Input:  $AR = \{rule, support, conf\}$  // association rules without updating confidence
          $S' = \{operational\ sequence\}$  // sequence data with corresponding state change
          $\epsilon$  // real minimum confidence

Output:  $AR = \{rule, support, confidence\}$ 

```

---

```

1: for all rules  $\in AR$  do
2:    $count = support$ 
3:   for  $s$  in  $S'$  do
4:      $pre = getAntecedentSequence(rule)$ 
5:     if ( $isSubsequence(pre, s)$ ) then  $count += 1$ 
6:   end for
7:    $update\_conf = support/count$ 
8:   if ( $confidence \geq \epsilon$ ) then  $confidence = update\_conf$ 
9:   else then  $filter(rule)$ 
10: end for

```

---

FIGURE 6: Confidence updating method.

**3.1. Description of the Data.** To ensure the credibility of the conclusions obtained from association rule mining, we extracted 20 batches of cold start-up data from the historical operating data of a commercial NPP. No operation errors or alarms are known to have occurred with these data. We set the start time of each batch of cold start-up data as the time of the first control rod action (the lifting order of the control rod is represented as  $I \rightarrow II \rightarrow III$ ), and the deadline time as the time when the reactor reaches stable power operation. The total sample size of the extracted data was 357595.

**3.2. Result of Data Preprocessing.** During the 3 Stages involved in a cold start-up, the operator is primarily concerned with the response of the neutron period of the source range channel in Stage 1, the changes in pressure in Stage 2, and the changes in temperature, pressure, and flow in Stage 3. We used the threshold definition method, the differential definition method, and the moving average definition method to define the source range channel-related state switching values, and the differential definition method to define the state switching values related to the pressure of the pressurizer, the primary loop temperature, and flow, depending on the distribution characteristics of the parameters. The detailed description of the defined state switching values is shown in Table 1.

Based on the generated state switching values, we reduced the data dimensionality to 272 using the feature selection algorithm proposed in [34]. The sequence segmentation method in Section 2.1 is then executed, where the maximum time interval is set to 100 sec, and the maximum sequence length is set to 30.

**3.3. Results and Analysis of the Association Rule.** In this study, the threshold  $\gamma$  of minimum support was set to 3 to remove rare association rules. Although setting a low threshold may affect the efficiency of the algorithm, it is meaningful to reveal the influence of the operational sequence on the NPP operation. For example, rules with long sequences often have low support, but they may have high confidence and can be

analyzed in comparison with subsequences with high support. The true minimum confidence  $\epsilon$  was set to 0.6 to remove weak association rules. Finally, a total of 164 original association rules are generated for further analysis, some of which are shown in Table 2.

All association rules are grouped according to the state switching value changes in the consequence sequence, and the rules are presented in the form of *operational sequence*  $\implies$  *state switching value changes*. After careful examination of all the association rules, four typical cases found are discussed by experts. It should be noted that the results and recommendations generated by the discussion are only applicable to the source of data collection.

**Case 1.** Close the emergency shut-down signal on the source range channel.

Following the operating instruction, the operator should close the emergency shut-down signal of the source range channel when the corresponding threshold is exceeded [35]. Rules 1–6 show that the operation of closing this signal has high support and confidence with the  $0 \rightarrow 1$  change of the state switching value related to the source range channel (point 2) defined by the moving average definition method. The maximum support of the relevant rule is 18, indicating that the phenomenon did not occur only 2 times out of 20 batches of cold start-up data. Furthermore, the rules show that the more frequently the operator withdraws the control rod in the period before performing the operation to close this signal, the lower the support and the higher the confidence in the  $0 \rightarrow 1$  change of the defined source range channel (point 2) with the moving average state switching value. In addition, in the rules related to this operation, although there are rules in which the state switches to  $1 \rightarrow 0$ , the operation is the last item in the antecedent sequence.

The above findings suggest that the operation of closing the emergency shut-down signal of the source range channel causes a local peak in the source range channel. This peak is influenced by the rate at which the control rod is withdrawn before the closing operation. The higher the rate, the more

TABLE 1: The detailed description of the defined state switching values.

Name of analog quantities	Ranking method	Definition method	$m$	$n$ (%)
Source range channel (point 1/point 2)	Ascending	Threshold definition	—	0.03
		Differential definition	10	
		Moving average definition		
Pressure of pressurizer	Descending	Differential definition	—	
Temperature of the primary loop				
Flow of the primary loop				

TABLE 2: Part of the association rules.

Number	Rule	Support	Confidence
1	{Close source range channel: 0→1} ⇒ {source range channel (point 2) of moving average: 0→1}	18	0.62
2	{Control rods of III withdrawal: 0→1, close source range channel: 0→1} ⇒ {source range channel (point 2) of moving average: 0→1}	15	0.83
3	{Control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, close source range channel: 0→1} ⇒ {source range channel (point 2) of moving average: 0→1}	7	0.875
4	{Control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, close source range channel: 0→1} ⇒ {source range channel (point 2) of moving average: 0→1}	5	1
5	{Control rods of III withdrawal: 0→1, close source range channel: 0→1} ⇒ {source range channel (point 2) of moving average: 1→0}	15	0.83
6	{Control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, control rods of III withdrawal: 0→1, close source range channel: 0→1} ⇒ {source range channel (point 2) of moving average: 1→0}	8	0.8
7	{Control rods of III withdrawal: 0→1, close source range channel: 0→1} ⇒ {source range channel (point 1) of moving average: 0→1}	5	0.625
8	{Control rods of II withdrawal: 0→1, control rod of II lower limit: 1→0} ⇒ {source range channel (point 2) of differential: 0→1}	6	0.67
9	{Control rods of I withdrawal: 0→1, control rods of I withdrawal: 0→1, control rods of II withdrawal: 0→1, control rod of II lower limit: 1→0} ⇒ {source range channel (point 2) of differential: 0→1}	4	1
10	{Control rods of III withdrawal: 0→1, control rod of III lower limit: 1→0} ⇒ {source range channel (point 2) of threshold: 0→1}	5	0.625
11	{Control rods of II withdrawal: 0→1, control rods of III withdrawal: 0→1, control rod of III lower limit: 1→0} ⇒ {source range channel (point 2) of threshold: 0→1}	4	0.67
12	{Control rods of III withdrawal: 0→1, control rod of III lower limit: 1→0} ⇒ {source range channel (point 1) of differential: 0→1}	3	1

likely the phenomenon is to occur. After a short period, the source range channel value gradually returns to normal, and this recovery process is unaffected by the subsequent withdrawal of the control rod. We make the following recommendations based on these rules.

- (i) The operator should carefully monitor the change in the source range channel value for a while before and after closing the emergency shut-down signal of the source range channel until the source range channel value stabilizes.
- (ii) Before closing the emergency shut-down signal of the source range channel, the operator should reduce the control rod withdrawal rate by reducing the frequency of withdrawal or the duration of a single

rod withdrawal to avoid greater fluctuations caused by this operation.

*Case 2.* Withdraw control rods from the lower limit position.

Before the cold start-up of the reactor, each control rod is located in the lower limit position. To achieve a cold start-up, multiple sets of control rods need to be withdrawn to the specified position. Rules 8–11 show that withdrawing the control rods of II and III from the lower limit position has a certain correlation with the change of 0 → 1 of the state switching values of the source range channel defined by the threshold definition method and the differential definition method, and the maximum support of the corresponding rules is 6.



According to prior knowledge, control rods I and II need to be withdrawn before control rods II and III are withdrawn from the lower limit position. The above findings indicate that when one set of control rods is in the lower limit position, the continuous operation of the control rod withdrawal with the last set of control rods will increase the change rate of the source range channel. This phenomenon is not conducive to the safe operation of the NPP. Accordingly, we propose the suggestion that when a set of control rods is in the lower limit position and needs to be withdrawn, the operator should avoid continuous withdrawal operations with the previous set of control rods, thus reducing the average withdrawal rate of multiple sets of control rods and preventing a large source range channel rate of change.

*Case 3.* The source range channels of points 1 and point 2 are not uniformly affected by the same operation.

Based on Case 1 results, it can be further found that the number of state switching values containing the source range channel definition of point 1 in the rule is small, and the support and confidence are low. This demonstrates that the nonsmooth variation in the source range channel due to the operations of the operator in the rules generated by the mining is mainly reflected in point 2. Therefore, operators should pay more attention to the source range channel of point 2 during the operation of the cold start-up.

*Case 4.* No rules related to the state switching values defined by the primary loop temperature and flow rate are generated.

It is worth noting that although the differential definition method was used to define the state switching values related to the pressure of the pressurizer, temperature, and flow of the primary loop, it did not produce valuable rules related to changes in these state switching values. According to the mining results of these historical operating data, if there is a large instantaneous change rate of the pressure of the pressurizer, temperature, and flow of the primary loop under the premise of obeying the operating instruction during the cold start-up, the operator can preliminarily determine that it is not caused by the operations of the operator. This is when the system may be abnormal or faulty, and the cause needs to be further identified.

## 4. Conclusions

Nowadays, NPPs have accumulated a large amount of historical operating data, from which it is valuable to reveal the impact of operational sequence on their operation. In this paper, a method based on association rule mining is proposed to analyze the operational sequence characteristics of operators and their impact on the NPPs operating. We verify the effectiveness of the proposed method using 20 batches of cold start-up historical operating data. The results show that the raw data can be converted into segmented sequence data sets according to the defined state switching value. 164 original association rules were obtained using the accompanying mining technique and its postprocessing solution. These rules reveal some valuable operational issues,

such as the effect of control rod action on the neutron period under specific conditions.

The advantage of the proposed method is that it can flexibly mine the association between the operational sequences of the operator and the operation phenomenon under study through the artificially defined state switching value. In particular, it is possible to mine operation phenomena that do not trigger an alarm but may harm the safe operation of the NPP. The rules mined can also be used to guide the operations of operators to avoid the recurrence of these phenomena. It is important to note that the results obtained based on association rule mining do not indicate a causal relationship between the operations and the operational phenomenon to which the rule corresponds, and further analysis of the rule is necessary. Nevertheless, this approach can be used to mine rule patterns as long as historical operating data for several batches (the same scenario) is available. The proposed method could be further explored for extended applications in the future. First, the proposed method can be improved for online data mining. Second, the training data of the operator is extracted, and the operations of the operator are evaluated by this method. Third, it is used to study the correlation between different operation phenomena of NPPs and explore the implied nonlinear relationship between different parameter variables.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] Y. Zou, Z. Xiao, L. Zhang, E. Zio, J. Liu, and H. Jia, "A data mining framework within the Chinese NPPs operating experience feedback system for identifying intrinsic correlations among human factors," *Annals of Nuclear Energy*, vol. 116, pp. 163–170, 2018.
- [2] L. Zheng, F. Jiao, P. Liu, J. Tong, and Z. Li, "Operating events in Chinese nuclear power plants: a preliminary analysis," *Advances in Intelligent Systems and Computing*, vol. 589, pp. 336–343, 2018.
- [3] R. Hamer, P. Waterson, and G. T. Jun, "Human factors and nuclear safety since 1970 – a critical review of the past, present and future," *Safety Science*, vol. 133, Article ID 105021, 2021.
- [4] M. Sethu, B. Kotla, D. Russell et al., "Application of artificial intelligence in detection and mitigation of human factor errors in nuclear power plants: a review," *Nuclear Technology*, vol. 209, no. 3, pp. 276–294, 2023.
- [5] M. Hamza and M. Diaconeasa, "Investigating the applicability of human reliability analysis methods during early design stages of non-light-water nuclear power plants," *Progress in Nuclear Energy*, vol. 161, Article ID 104716, 2023.
- [6] J. Liu, Y. Zou, W. Wang et al., "A study on assigning performance shaping factors of the SPAR-H method for

- adequacy human reliability analysis of nuclear power plants,” *International Journal of Industrial Ergonomics*, vol. 81, Article ID 103051, 2021.
- [7] J. Park, R. L. Boring, T. A. Ulrich et al., “A framework to collect human reliability analysis data for nuclear power plants using a simplified simulator and student operators,” *Reliability Engineering and System Safety*, vol. 221, Article ID 108326, 2022.
  - [8] M. Porthin, M. Liinasuo, and T. Kling, “Effects of digitalization of nuclear power plant control rooms on human reliability analysis – a review,” *Reliability Engineering and System Safety*, vol. 194, Article ID 106415, 2020.
  - [9] J. Park, Y. Kim, and W. Jung, “Calculating nominal human error probabilities from the operation experience of domestic nuclear power plants,” *Reliability Engineering and System Safety*, vol. 170, pp. 215–225, 2018.
  - [10] I. Jang, Y. Kim, and J. Park, “Investigating the effect of task complexity on the occurrence of human errors observed in a nuclear power plant full-scope simulator,” *Reliability Engineering and System Safety*, vol. 214, Article ID 107704, 2021.
  - [11] C. Oh and J. Lee, “Real time nuclear power plant operating state cognitive algorithm development using dynamic Bayesian network,” *Reliability Engineering and System Safety*, vol. 198, Article ID 106879, 2020.
  - [12] Y. Zou, L. Zhang, L. Dai, P. Li, and T. Qing, “Human reliability analysis for digitized nuclear power plants: case study on the LingAo II nuclear power plant,” *Nuclear Engineering and Technology*, vol. 49, no. 2, pp. 335–341, 2017.
  - [13] G. Hu, T. Zhou, and Q. Liu, “Data-driven machine learning for fault detection and diagnosis in nuclear power plants: a review,” *Frontiers in Energy Research*, vol. 9, 2021.
  - [14] Q. Hong, M. Jun, W. Bo et al., “Application of Data-Driven technology in nuclear Engineering: prediction, classification and design optimization,” *Annals of Nuclear Energy*, vol. 194, Article ID 110089, 2023.
  - [15] S. K. Solanki and J. Patel, “A survey on association rule mining,” in *Proceedings of the 2015 Fifth International Conference on Advanced Computing and Communication Technologies*, pp. 212–216, Enathi, India, December 2015.
  - [16] J. Jiang, L. Zhang, Y. Wang, K. Zhang, D. Yang, and W. He, “Association rules analysis of human factor events based on statistics method in digital nuclear power plant,” *Safety Science*, vol. 49, no. 6, pp. 946–950, 2011.
  - [17] Y. He, P. Gu, W. Xi, and T. Bai, “Research on status monitoring reliability prediction method of digital control system for nuclear power plant,” *Atomic Energy Science and Technology*, vol. 51, pp. 2338–2343, 2017.
  - [18] Y. Xu, *Research and Implementation of Fault Dianosis and Prediction System for Nuclear Power Equipment*, Huazhong University of Science and Technology, Wuhan, China, 2019.
  - [19] Y. Li, T. Li, and H. Liu, “Recent advances in feature selection and its applications,” *Knowledge and Information Systems*, vol. 53, no. 3, pp. 551–577, 2017.
  - [20] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, “Dimensionality reduction for fast similarity search in large time series databases,” *Knowledge and Information Systems*, vol. 3, pp. 263–286, 2001.
  - [21] R. Agrawal, K. Lin, H. Sawhney, and K. Shim, “Fast similarity search in the presence of noise, scaling, and translation in time-series databases,” in *Proceedings of the 21th International Conference on Very Large Data Bases*, pp. 490–501, San Francisco, CA, USA, September 1995.
  - [22] D. Rafiei and A. Mendelzon, “Efficient retrieval of similar time sequences using DFT,” *Information Organization and Databases*, vol. 579, pp. 75–89, 2000.
  - [23] J. Lin, E. Keogh, L. Wei, and S. Lonardi, “Experiencing SAX: a novel symbolic representation of time series,” *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 107–144, 2007.
  - [24] D. Bao and Z. Yang, “Intelligent stock trading system by turning point confirming and probabilistic reasoning,” *Expert Systems with Applications*, vol. 34, no. 1, pp. 620–627, 2008.
  - [25] K. Pratt, H. Corporation, and E. Fink, “Search for patterns in compressed time series,” *International Journal of Image and Graphics*, vol. 2, 2001.
  - [26] R. Agrawal and R. Srikant, “Mining sequential patterns,” in *Proceedings of the Eleventh International Conference on Data Engineering*, pp. 3–14, Taipei, Taiwan, March 1995.
  - [27] J. Han, H. Cheng, D. Xin, and X. Yan, “Frequent pattern mining: current status and future directions,” *Data Mining and Knowledge Discovery*, vol. 15, no. 1, pp. 55–86, 2007.
  - [28] J. Pei, J. Han, B. Mortazavi-Asl, and H. Pinto, “PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth,” in *Proceedings 17th International Conference on Data Engineering*, pp. 215–224, Heidelberg, Germany, April 2001.
  - [29] M. Zaki, “SPADE: an efficient algorithm for mining frequent sequences,” *Machine Learning*, vol. 42, no. 1/2, pp. 31–60, 2001.
  - [30] M. Zaki, “Scalable algorithms for association mining,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, 2000.
  - [31] R. Yu, Z. Chen, H. Yang, Y. Chen, K. Xiao, and Y. Liu, “Study on automatic control method for clod start-up of nuclear power plant,” *Hedongli Gongcheng/Nuclear Power Engineering*, vol. 37, no. 1, pp. 62–66, 2016.
  - [32] J. Chen J, *Daya Bay Nuclear Power Plant System and Operation*, Atomic Energy Press, Gujarat, India, 1994.
  - [33] X. Qing, K. Xiao, K. Huang, G. Chen, Y. Li, and Z. Chen, “Simulated analysis of pressure control and overpressure problem of reactor cold start-up with nuclear heating,” *He Dong Li Gong Cheng/Nuclear Power Engineering*, vol. 41, pp. 84–88, 2020.
  - [34] Y. He, H. Yu, R. Yu et al., “A correlation-based feature selection algorithm for operating data of nuclear power plants,” *Science and Technology of Nuclear Installations*, vol. 2021, 2021.
  - [35] Guangdong Nuclear Power Training Center, *System and Equipment for 900 MW PWR Nuclear Power Plant*, Atomic Energy Press, Gujarat, India, 2005.