



**International Journal of Innovation and Sustainable Development**

ISSN online: 1740-8830 - ISSN print: 1740-8822

<https://www.inderscience.com/ijisd>

---

**Risk warning method for the whole process of production project based on multi-source data mining**

Ya Zhou, Mingjie Zhang, Shaohuan Cheng, Xinli Luo, Jiefeng Wen, Jinlei Hu

**DOI:** [10.1504/IJISD.2025.10067825](https://doi.org/10.1504/IJISD.2025.10067825)

**Article History:**

|                   |                  |
|-------------------|------------------|
| Received:         | 18 August 2022   |
| Last revised:     | 08 October 2022  |
| Accepted:         | 31 October 2022  |
| Published online: | 02 December 2024 |

---

## Risk warning method for the whole process of production project based on multi-source data mining

---

Ya Zhou, Mingjie Zhang, Shaohuan Cheng,  
Xinli Luo, Jiefeng Wen and Jinlei Hu\*

Qingyuan Power Supply Bureau of Guangdong Power Grid Co., Ltd.,  
Qingyuan, 511500, China

Email: st.cloud@163.com

Email: 13926672324@qy.gd.csg.cn

Email: 345654669@qq.com

Email: 234242236@qq.com

Email: 367206012@qq.com

Email: hjinlei@mls.sinanet.com

\*Corresponding author

**Abstract:** Aiming at the problem that the correlation coefficient between the final screening early warning indicators is small in the whole process of risk index screening of production projects, a risk early warning method for the whole process of production projects based on multi-source data mining is proposed. The multi-source data mining technology is used to extract the potential hidden characteristics of risk indicators in the whole process of production projects, and the hidden characteristic relationship between multi-risk indicators is obtained, and the risk management index system of the whole process of production projects is established. Risk warning of the whole process of production projects. The test results show that the proposed risk early warning method has a larger cross-correlation coefficient calculated in each index group, which verifies that the early warning method has high application reliability, can obtain more accurate data, and improve the early warning accuracy.

**Keywords:** multi-source data mining; production project; whole process risk; risk early warning; K-means cluster analysis algorithm; index correlation characteristics.

**Reference** to this paper should be made as follows: Zhou, Y., Zhang, M., Cheng, S., Luo, X., Wen, J. and Hu, J. (2025) 'Risk warning method for the whole process of production project based on multi-source data mining', *Int. J. Innovation and Sustainable Development*, Vol. 19, No. 1, pp.94–106.

**Biographical notes:** Ya Zhou obtained his Bachelor's degree in Electrical Engineering and Automation from Xi'an Jiaotong University in 2004. He is currently an Engineer of Qingyuan Power Supply Bureau of Guangdong power grid company, mainly engaged in distribution network automation, production project management, etc.

Mingjie Zhang graduated from Guangdong University of Technology with a Master's degree in Electrical Engineering, is currently a Senior Engineer of Qingyuan Power Supply Bureau of Guangdong Power Grid Co., Ltd., mainly engaged in distribution network operation and reliability management.

Shaohuan Cheng obtained his Bachelor's degree in Computer and its Application from Northeast Electric Power University. He is currently a Senior Engineer of Qingyuan Power Supply Bureau of Guangdong Power Grid Co., Ltd., mainly engaged in distribution network planning and distribution network informatisation.

Xinli Luo received his Bachelor's degree in Electrical Engineering and Automation from Hunan University of Science and Technology. Currently, he is an Engineer of Qingyuan Power Supply Bureau of Guangdong Power Grid Company, mainly engaged in power transformation operation and technical transformation.

Jiefeng Wen received his Bachelor's degree in Electrical Engineering and Automation from Guangdong Ocean University. He is currently an Engineer of Qingyuan Fresh Power Supply Bureau of Guangdong Power Grid Co., Ltd., mainly engaged in distribution network planning.

Jinlei Hu is a PhD candidate in High Voltage Technology of South China University of Technology, is currently a Senior Engineer of Qingyuan Power Supply Bureau of Guangdong Power Grid Company, mainly engaged in relay protection of power system, power market, power grid operation, etc.

---

## **1 Introduction**

Production projects in some special industries have certain risks, such as power production. There are more electrical facilities and high voltage equipment in the electric power production environment, more dangerous goods, and most of the electrical equipment is always in operation, resulting in a high incidence of accidents and serious consequences in the electric power production process (Chen et al., 2020). In order to ensure the safety of life and property in the production process, real-time risk warning for the entire process of production projects is the necessary condition and important guarantee for modern production safety (Zhang and Li, 2021; Lei et al., 2021). The core issue of risk early warning is the accuracy and time of early warning. The higher the accuracy, the better the warning effect. The shorter the warning time, the more quickly the fault point can be found, the potential hazards of electrical equipment can be solved in time, and serious accidents can be prevented.

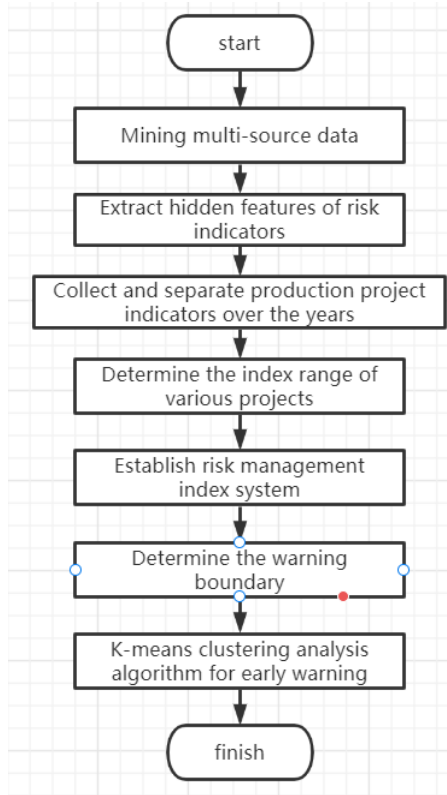
Li et al. (2020) uses the risk matrix method to evaluate the risk of hydropower plant production projects, and divides the existing risk into four different levels. The research has a reference and reference for the risk analysis and safety management of hydropower stations. Long et al. (2020) combined K-means clustering, FOA and RF algorithms to construct a real-time early warning model for coal and gas outburst based on data mining, which realised real-time and accurate early warning of coal and gas outbursts at working face. The above methods provide a reference for risk early warning in related fields, but there are individual differences between different production industries. When there are program sets and project portfolios on top of projects, the actual comprehensive judgement and early warning capabilities of managers are often difficult to manage and control risks. High standard requirements.

Based on the existing research results, this paper proposes a risk early warning method for the whole process of production projects based on multi-source data mining. Precise early warning.

## 2 Design of risk warning method for the whole process of production project

The design process of the risk early warning method for the entire process of a production project is shown in Figure 1.

**Figure 1** Design process of risk early warning method in the whole process of production project (see online version for colours)



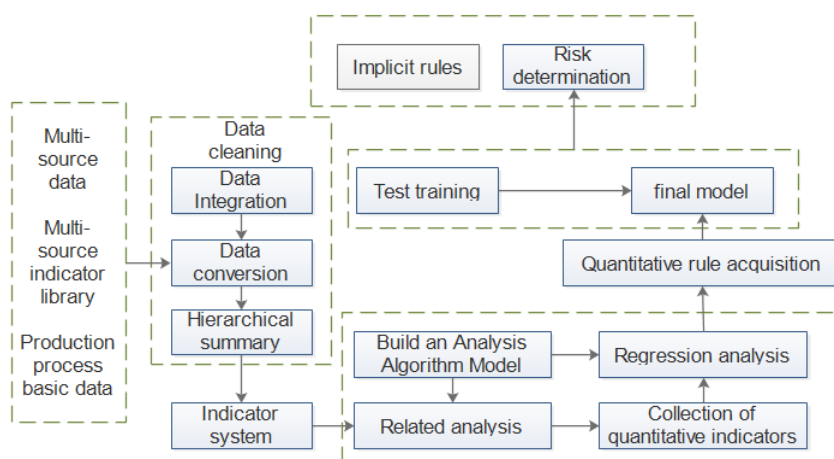
According to the design process of risk early warning method in the whole process of production project in Figure 1, the detailed design method is as follows.

### 2.1 Extraction of hidden features of risk indicators in the whole process of production projects

Through the mining of multi-source data in the whole process of risk warning of production projects, the implicit relationship and characteristics existing between two seemingly unrelated indicators can be analysed. In the whole process risk early warning

analysis of production projects based on multi-source data mining technology, the correlation between risk early warning indicators also needs to excavate the internal implicit relationship characteristics (Ren et al., 2020). The whole process risk warning of production projects can measure the change trend between the actual production state and the theoretical risk warning limit. Therefore, during the risk warning period of the whole process, it is necessary to extract the risk indicators that have a greater impact on production from real-time production data, and analyse the risk indicators. Analyse internal characteristics based on changes in indicator data. In the overall process of this data analysis, a schematic diagram of the multi-source data mining process involved is shown in Figure 2.

**Figure 2** Diagram of multi-source data mining process (see online version for colours)



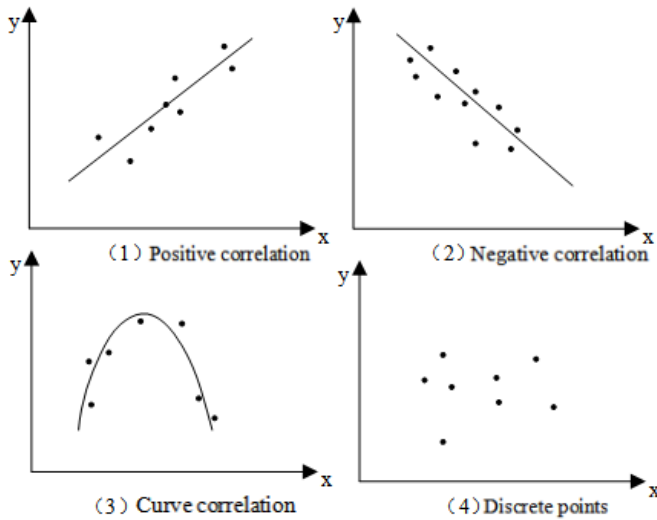
Sort out the system process and operating standard requirements involved in the whole process management of production projects; investigate the source, content and storage format of existing information in the production project management system; collect and analyse the main problems occurred in the process of production projects over the years, and determine the main risk points (Wang et al., 2020; Zhang et al., 2021). After sorting out the audit problems of production projects over the years, classify the projects and problems, and look for the common features and internal connections in similar projects and similar problems. Therefore, in the data collection stage, according to the type and characteristics of the actual production project, the quality of the collected data is evaluated, and the result of the evaluation is used as the final guiding direction to integrate the accuracy and comprehensiveness of the whole process data of the production project. The hidden characteristics of risk indicators must be easy to obtain and have applicability.

In general, the entire production process of the production project is regarded as a data cycle, so that the obtained index data is larger, more complete, and has obvious change rules, which has important value in the analysis process. In the process of risk warning of this process, the source of the data should be reliable. According to the actual situation of the specific production project, try to use a combination of various data collection methods to obtain process indicators (Lei et al., 2021). If there are missing data in the selected indicators, then It is necessary to analyse the missing link of the

production project data, and ask the corresponding professional department to be responsible for manual input to ensure that the data is complete and accurate, and store it in a structured way (Huang and Feng, 2020; Shu, 2021).

The process of common production projects, deviations in production units or index parameters may cause the system to transition from steady state to non-steady state. Obtain relevant information on job site risk factors and system state parameters, comprehensively analyse and predict the risk factors and mutation trends of the production process, and reasonably predict the accident probability and the impact range of accident consequences based on historical accident data. After the data collection is completed, the data needs to be cleaned, and the discrete data and the data that are out of cycle accuracy are filtered and eliminated. Among the hidden features between risk indicators, the existence form of each data is in the form of scatter. Each data scatter in the same coordinate system will show different relationships, as shown in Figure 3.

**Figure 3** Risk indicator recessive feature relation diagram



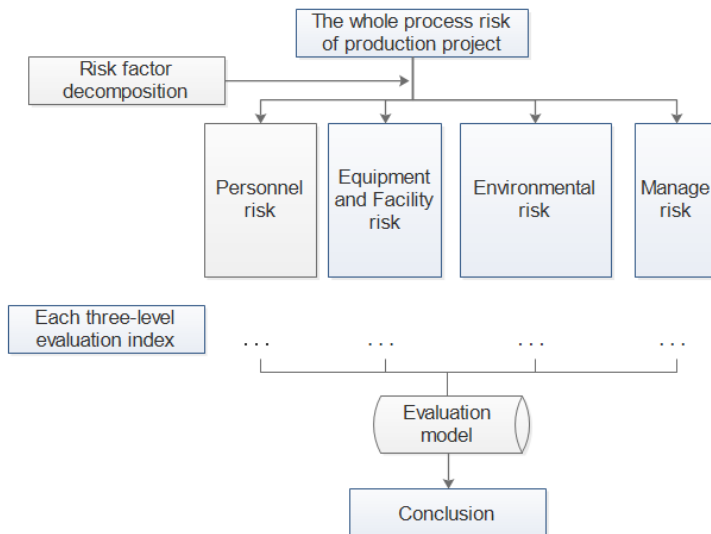
Relationships shown in Figure 3, the correlation coefficient can be used to describe the degree of closeness between the indicators. So far, the extraction of the hidden features of the risk indicators in the whole process of the production project has been completed.

## 2.2 Risk management indicator system for the whole process of production projects

Based on the hidden feature information of risk indicators extracted from the whole process of production projects, an index system for risk management in the whole process of production projects is established. According to the requirements of the production project management system and operation instructions of the network and the provincial company, establish the process indicators of each link of the production project, collect and separate the production project indicators of the past years from the existing database, use the big data method to clean and analyse the data, and determine various projects index range. In the implementation of the whole process of production projects,

it is very difficult to carry out comprehensive risk assessment for projects with complex processes, great difficulties and many influencing factors. Therefore, it is necessary to establish a risk management index system for the whole process of production projects (Zhang et al., 2020; Wang et al., 2020). In the process of system construction, the principles of scientificity, systematicness and rationality should be followed. On this basis, determine the index factors, and extract the hidden characteristics of risk indicators in the whole process of production projects based on multi-source data mining obtained above, and the obtained evaluation index system is shown in Figure 4.

**Figure 4** Risk assessment index system for the whole process of production projects (see online version for colours)



Risk assessment index system is an important tool for risk assessment in the whole production process. The risk assessment index system needs to have a high degree of matching with the risks of the production process, so as to ensure the search, analysis and prediction of dangerous and harmful factors in the production process. Carry out objective description and analysis, and predict the possible harmful consequences of the above indicators, and analyse the severity and final impact of the consequences. In the process of establishing the risk assessment index system, the characteristics of the production project and the difficulties and points of each production link should be considered. According to the actual production project situation, the matching risk index hidden characteristics are selected as evaluation, and on this basis, a reasonable production plan and risk countermeasures are customised. So far, the establishment of the risk management indicator system for the whole process of the production project has been completed.

### 2.3 Optimised risk identification and early warning algorithm

Based on the characteristics of production projects in the risk management index system, customise and optimise the targeted warning algorithm. The main purpose of the typical production project risk early warning method in the whole process is to describe the

change trend between the production state and the preset early warning limit. Therefore, in the early warning process, the early warning limit must be determined first. The risk warning boundary and the warning classification can effectively describe the severity of the accident and directly reflect the risk status. According to different production projects, the corresponding boundaries are formulated, which are generally divided into 2–5 early warning boundaries. After completing the setting of the warning limit, it is necessary to calculate the index value of the whole process of the real production project. The risk identification and early warning algorithm used in this paper is the K-means cluster analysis algorithm (Li and Wang, 2020; Liu et al., 2020). The principle of this algorithm is to calculate the distance between the sample and the cluster center to distinguish the category of the sample. The accident index data obtained through the above calculations are combined, and the attribute vectors contained in each data are analysed. Set an expression that sums two of these  $Y$  data items  $X$ :

$$\begin{cases} X = (x_1, x_2, \dots, x_n) \\ Y = (y_1, y_2, \dots, y_n) \end{cases} \quad (1)$$

Randomly extract a number of data from the set of formula (1), and use it as the centre point of the initial clustering of each category, calculate the distance between the data around the centre point and the centre point, and distribute them according to the principle of distance to ensure that each The similarity in a class of data around the cluster center point is at a relatively high level. In the process of risk warning, the following functional relationship exists between the risk response and the frequency response function:

$$H(\omega) = \frac{\alpha}{\beta} XY \quad (2)$$

Among them,  $H(\omega)$  represents the frequency response function, represents the  $\alpha$  system response  $\beta$  in the process of risk early warning, and represents the excitation coefficient. According to the objective relationship that exists between period  $\omega$  and frequency  $f$ :

$$\omega = 2\pi f \quad (3)$$

From a macro point of view, the whole-process risk warning of a production project belongs to the category of changes in degrees of freedom, and its control equation can be expressed as:

$$y(f) = x_n(m + c) + kx(f) \quad (4)$$

Among them, it represents the frequency change function of the  $x(f)$  system response in the  $y(f)$  risk warning process, the frequency change function of the excitation coefficient, the  $m$  risk characteristic coefficient, the  $c$  determination coefficient of the severity of the risk consequences, and the application coefficient of the  $k$  K-means cluster analysis algorithm. Under the above conditions, the corresponding risk identification and early warning frequency response function can be obtained as follows:



The power production project in Figure 5, the production equipment and the production process share safe and reliable equipment, instruments and raw materials, on this basis, the safe implementation of the production project can be ensured. However, due to the characteristics of power production raw materials and products, they are all flammable raw materials, which can cause combustion and explosion under high temperature and open flame conditions. During the production process, the electrical equipment runs under load for a long time, and the transformer or cable will be abnormal and cause fire and explosion, etc. The fire and explosion hazard levels that may occur in the production project are shown in Table 1.

**Table 1** Fire and explosion hazard level of power production project installations

| <i>Workplace</i>             | <i>Risk category</i> | <i>Hazard level</i>      |
|------------------------------|----------------------|--------------------------|
| High pressure heat exchanger | Fire and explosion   | Class I (High risk)      |
| High pressure separator      | Fire and explosion   | Class I (High risk)      |
| Main stripping tower         | Fire and explosion   | Class II (Moderate risk) |
| Recycle reaction tower       | Fire and explosion   | Class I (High risk)      |
| Stator Contactor             | Fire and explosion   | Class II (Moderate risk) |

In the performance test, the risk early warning indicators in the production process of the power device are selected as the basis, and the historical operation data of the device are collected for risk research. The data collected in the past three years from January 2015 to December 2017 are shown in Table 2.

**Table 2** Risk management index data of some operation of power installations

| <i>Risk management indicators</i>                     | <i>Year of<br/>2015</i> | <i>Year of<br/>2016</i> | <i>Year of<br/>2017</i> |
|-------------------------------------------------------|-------------------------|-------------------------|-------------------------|
| Process inherent hazards                              | 1.6                     | 3.3                     | 3.6                     |
| Equipment inherent hazards                            | 6.8                     | 6.6                     | 6.5                     |
| Inherent hazard of raw materials                      | 5.5                     | 5.5                     | 5.8                     |
| Production, storage and transportation integrity rate | 64.2                    | 68.1                    | 78.9                    |
| Production response mutation rate                     | 8.6                     | 2.1                     | 7.8                     |
| Operation error rate                                  | 3.5                     | 3.5                     | 1.5                     |
| Violation rate                                        | 9.4                     | 6.5                     | 1.41                    |
| Difficulty in rectifying hidden dangers               | 5.1                     | 7.5                     | 2.1                     |
| Equipment maintenance rate                            | 2.1                     | 1.2                     | 5.2                     |
| Personnel self-rescue ability                         | 4.5                     | 23.2                    | 2.2                     |
| Emergency response and rescue capability              | 23.85                   | 6.1                     | 6.2                     |
| Time required to resume normal production             | 85.5                    | 72.1                    | 96.2                    |
| Production site residual risk                         | 1.5                     | 85.5                    | 5.1                     |
| Technical security management                         | 2.85                    | 3.9                     | 2.1                     |
| Device security management                            | 8.5                     | 6.8                     | 3.2                     |
| Fire safety equipment                                 | 2.2                     | 9.2                     | 56.5                    |

**Table 2** Risk management index data of some operation of power installations (continued)

| <i>Risk management indicators</i>    | <i>Year of<br/>2015</i> | <i>Year of<br/>2016</i> | <i>Year of<br/>2017</i> |
|--------------------------------------|-------------------------|-------------------------|-------------------------|
| Cognition evaluation of risk ability | 43.5                    | 41.2                    | 45.5                    |
| Hazard factor coupling               | 15.6                    | 12.2                    | 12.2                    |
| Hazard exposure time                 | 4.5                     | 2.2                     | 2.2                     |
| Hazard investigation time            | 1.5                     | 3.23                    | 5.2                     |
| Production staff working density     | 83.6                    | 85.2                    | 75.5                    |

The data shown in Table 2, under a certain production environment, the designed risk early warning method for the whole process of production projects based on multi-source data mining and the currently common risk early warning method based on principal component analysis are the goals. The cross-correlation of risk indicators of different early warning methods is calculated to verify the performance of risk early warning methods.

### 3.1.2 Comparison and analysis of experimental results

The cross-correlation function between the operation risk management indicators of the power plant, the average value of the calculation results of any two indicators is obtained. Assuming that any two of the indicators are respectively  $P$  and  $W$ , then the process data sample values of these two indicators  $P$  are respectively  $P_n$  and  $W_n$ , and the formula for  $W$  calculating  $P$  the cross-correlation coefficient between the sums is as follows:

$$\sigma_{PW} = \frac{\sum_{n=1}^n (P_n - \bar{P})(W_n - \bar{W})}{\sqrt{\sum_{n=1}^n (P_n - \bar{P})^2} \sqrt{\sum_{n=1}^n (W_n - \bar{W})^2}} \quad (7)$$

In the formula,  $\bar{P}$  and  $\bar{W}$  represent the average value of the calculation results of these two indicators respectively. The larger the calculation results of the correlation number between indicators, the greater the correlation between indicators, and the more accurate the data of the whole process of production projects obtained through each indicator.

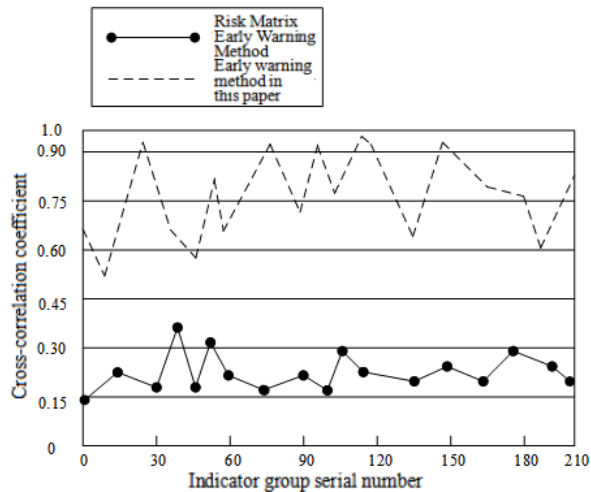
### 3.2. Comparison and analysis of experimental results

Under the formula (7) calculation method, the cross correlation coefficient transformation of the early warning method and the risk early warning method based on the risk matrix in each risk indicator is obtained through statistics, as shown in Figure 6.

It can be seen from Figure 6 that the correlation number calculated by the risk early warning method based on the risk matrix in different indicator groups fluctuates between 0.15 and 0.4, and the overall correlation coefficient is relatively low; The correlation number of the risk early warning methods for the whole process of production projects based on multi-source data mining designed in this paper is higher than 0.5, and the maximum is 0.95. It shows that the experimental results obtained by using the early warning method in this paper are closer to the actual early warning results, which indirectly verifies the reliability of the early warning method in this paper, can obtain

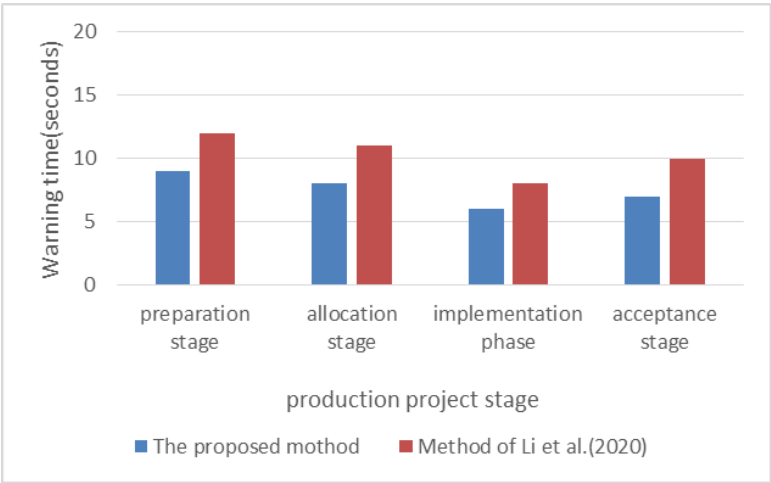
more accurate data, and improve the final early warning accuracy. This is because the early warning method in this paper adopts the K-means clustering analysis algorithm, which can combine the accident index data and analyse the attribute vectors contained in each data, so that the experimental results are closer to the actual results. Therefore, in practical application, the indicator data forewarned by using this method is easier to check and enhance the reliability of the final warning results.

**Figure 6** Experiment results of risk index correlation



The pre-warning time is an index to directly judge the pros and cons of the pre-warning method. Under the same conditions of the power plant operating environment, the proposed method and the method in Li et al. (2020) are selected to carry out the comparative experiment of the pre-warning time. The comparison results of early warning time are shown in Figure 7.

**Figure 7** Early warning time of different methods (see online version for colours)



It can be seen from Figure 7 that the minimum warning time required by the risk matrix warning method is 9 s, and the maximum warning time is 12 s; The warning time of the warning method in this paper is as short as 5 s and as long as 9 s, both of which are not more than 10 s. It shows that the early warning method in this paper is efficient and can send early warning signals in time to avoid risks and accidents caused by many indicators throughout the production project. This is because the method in this paper extracts the implicit characteristics of the risk indicators in the whole process of the production project, according to which the fault location can be found more accurately, the early warning speed can be accelerated, the large early warning error caused by the long warning time can be reduced, and the early warning signal can be sent in a timely manner to reduce the occurrence of faults.

To sum up, the risk indicators set by the risk early warning method for the whole process of production projects designed in this paper have strong correlation, high reliability and prediction efficiency, and better early warning performance.

## 4 Conclusion

Carry out research on the risk warning of the whole process of production projects based on multi-source data mining, use the existing production project data information to analyse the main risks of the production project process over the years, design and verify the project process risk judgement and warning rules, and realise the multi-source information of the project process. Automatic tracking and comparison, risk early warning, and auxiliary decision-making functions. This research can intelligently process a large number of business information stock documents generated in the whole process of production projects, identify potential risks, assist managers to manage production projects more comprehensively and scientifically, and reduce irregularities in the process of production projects.

## References

- Chen, C., Wang, X., Jiang, T., Xu, J. and Wei, L. (2020) 'Construction of influenza virus prevention and control prediction and early warning platform based on multi-source heterogeneous big data mining', *China Biotechnology*, Vol. 40, No. Z1, pp.109–115.
- Huang, D. and Feng, T. (2020) 'Research on social stability risk early warning of super-large engineering projects based on multiple conflict of interests', *Science and Technology Management Research*, Vol. 40, No. 15, pp.224–230.
- Lei, B., Xu, M., Deng, J., Gao, D.G. and Peng, P.(2021) 'Development and application of risk warning system for MAH production line based on ADR report data', *China Pharmacy*, Vol. 32, No. 08, pp.904–910.
- Li, G. and Wang, X. (2020) 'Research on credit risk early warning of, P.2P institutions based on GRBF neural network under the background of big data', *Practice and Understanding of Mathematics*, Vol. 50, No. 17, pp.35–43.
- Li, Y., Liu, Y., Hu, D., Hu, K.X. and Chen, J.X.(2020) 'Research on risk classification assessment method of hydropower station project based on risk matrix: taking Wudongde hydropower station as an example', *China Safety Production Science and Technology*, Vol. 16, No. 01, pp.130–134.

- Liu, J., Li, Z. and Liu, H. (2020) 'Research on the monitoring and early warning of extreme risk spillovers between China's financial markets: Implementation based on the MVMQ-CAViaR method', *Economics and Management Research*, Vol. 41, No. 02, pp.19–29.
- Long, N., Yuan, M., Wang, G., Hu, J. and Du, J.P. (2020) 'Research on real-time early warning of coal and gas outburst based on data mining', *China Mining Industry*, Vol. 29, No. 11, pp.88–93+99.
- Ren, W., Xu, Z., Song, X., Feng, J. and Zhen, J.H. (2020) 'Research on the operation characteristics monitoring and evaluation method of digital metering system of distribution network based on data mining', *High Voltage Electrical Appliances*, Vol. 56, No. 08, pp.183–191.
- Shu, L. (2021) 'Cloud-side integrated coal mine safety production risk monitoring and early warning platform based on big data', *Coal Mine Safety*, Vol. 52, No. 05, pp.144–148.
- Wang, H., Wang, X., Wang, Z., Luo, J.F. and Qin, D. (2020) 'Cluster analysis of risky driving behavior in dangerous goods transportation based on data mining', *Transportation System Engineering and Information*, Vol. 20, No. 01, pp.183–189.
- Wang, T., Qu, L., Ni, L., He, P.K. and Peng, G. (2020) 'Early warning of karst tunnel collapse risk based on uncertain measure theory', *Journal of Xi'an University of Science and Technology*, Vol. 40, No. 02, pp.284–291.
- Zhang, C., Zhao, M. and Guan, J. (2021) 'Early warning model of financial risk portfolio of listed manufacturing companies based on lasso + SVM', *Practice and Understanding of Mathematics*, Vol. 51, No. 05, pp.1–12.
- Zhang, K. and Li, J. (2021) 'Multi-objective control simulation of software operation risk based on attribute checking', *Computer Simulation*, Vol. 38, No. 09, pp.306–309+385.
- Zhang, Q., Zhang, Y. and Ding, R. (2020) 'Macroscopic safety risk early warning of regional special equipment based on the perspective of supervision', *Journal of Safety and Environment*, Vol. 20, No. 02, pp.397–405.