ISSN: 2664-9640

Research on Risk Warning in Power Grid Assets based on Association Rule Mining Algorithm

Wenlong Xu*

School of Mathematics and Physics, Anhui Jianzhu University, Hefei 230601, China

*Corresponding Author

Abstract

Aiming at the low efficiency of the data risk warning model in the grid asset system, the risk warning model of association rule mining algorithm is proposed. Meanwhile, the time series is introduced into the association rule mining algorithm, which utilizes the temporal features to further ensure the timeliness of the risk warning. The simulation results show that the recall rates of data leakage, financial abnormal data, and data tampering are 99.09%, 97.5%, and 100.00%, respectively, and the accuracy rates are 99.91%, 82.98%, and 80.00%, respectively, with an overall classification accuracy rate of 99.04%. The overall classification accuracy rate is 99.04%. The accuracy rate of risk warning for data leakage is relatively high. The larger the data and the longer the waiting time of the grid assets will have a certain impact on the detection efficiency. Moreover, the average value of the warning delay of the machine learning algorithm and the extreme learning machine algorithm for grid asset data leakage, asset financial anomaly data and grid asset data tampering is more than 1200 ms, which is twice as much as that of the association rule mining algorithm. Therefore, the proposed association rule mining algorithm has a better risk warning performance with the advantages of low latency and short warning waiting time.

Keywords

Association Rule Mining; Grid Assets; Risk Warning; Waiting Time; Recall.

1. Introduction

Power grid asset risk early warning faces challenges of uncertainty and complexity. Within the context of big data, conducting accurate and effective power grid asset risk early warning is crucial for the sustainable operation of power enterprises. The evolution of power grid asset risk constitutes a dynamic process [1]. Traditional risk early warning models fail to capture the dynamic variations of risk factors inherent in this process [2-4]. Consequently, leveraging artificial intelligence algorithms within the new generation of information technology environments is essential for continuously enhancing the accuracy of power grid asset risk early warning models.

Yang et al. [5] investigated real-time enterprise asset risk monitoring based on cluster analysis, utilizing a weighted centroid algorithm to improve asset management efficiency and mitigate the risk of malicious data tampering. Their experimental study demonstrated that this method contributes to resolving asset data risk issues. Chen et al. [6] proposed the use of reinforcement learning algorithms for asset data risk detection. This algorithm can efficiently mine variable relationships within large databases in a short timeframe, thereby enhancing asset data risk detection capabilities. Huang [7] established a financial asset risk early warning model based on sentiment analysis and processing of information collected from the internet. This model exhibits a relatively small deviation between predicted financial risks based on big data and actual risks. However, existing algorithm models often neglect critical factors such as early

ISSN: 2664-9640

warning accuracy and latency, which significantly impact the efficiency of power grid asset risk early warning [8].

To address these limitations, this paper proposes the application of an association rule mining algorithm for power grid asset risk early warning analysis. Firstly, acknowledging the complexity of power grid asset systems, time series analysis is employed to investigate the temporal characteristics of power grid asset data, facilitating an in-depth study of the dynamic changes within the asset data. Subsequently, the association rule mining algorithm is utilized to identify frequently occurring fuzzy itemsets and determine pertinent association rules. Building upon this foundation, an asset risk early warning model based on the association rule mining algorithm is established. The model's risk early warning performance is comprehensively evaluated using key parameters including risk warning waiting time, warning recall rate, accuracy, and warning latency.

2. Risk Early Warning Model based on Association Rule Mining Algorithm

2.1. Apriori Algorithm Model

Association rule mining algorithms can identify correlations between different items within the same event.[9] Operating within the support-confidence framework, the Apriori algorithm model iteratively generates a structure containing multiple frequencies, characterized by an anti-monotonicity property.[10] Given a dataset containing d distinct items, the algorithm simultaneously generates frequent itemsets and association rules R:

$$R = \sum_{k=1}^{d-1} [\binom{d}{k} \times \binom{d-k}{j}] = 3^d - 2^{d+1} + 1 \tag{1}$$

If an itemset belongs to a frequent pattern, then all its subsets must also be frequent. This property is termed the anti-monotonicity of frequent patterns [11]. Conversely, if an item is identified as infrequent, the entire subgraph containing that item can be immediately pruned [12]. During rule generation, the support metric represents the probability of occurrence of power grid asset itemsets A and B in the database, signifying statistical significance.

2.2. Time Series Data Mining

To effectively enhance the efficiency of risk early warning for power grid assets, time series analysis is integrated into the association rule mining algorithm. Leveraging temporal characteristics helps further ensure the timeliness of risk warnings [13]. The inherent time-series properties of power grid asset systems are utilized for risk data mining [14]. Let Y represent the power grid asset time series, which can be expressed as:

$$Y = f(T, S, C, e) \tag{2}$$

where: T denotes the trend-cycle component, indicating the long-term, underlying pattern of the risk warning value steadily increasing, decreasing, or stabilizing over a specific period. S represents the seasonal variation, signifying regular, recurring fluctuations in risk warnings within specific periods (e.g., seasons). C is the cyclical component, reflecting long-term, non-seasonal cyclical fluctuations in the risk warning value. e represents the random error (irregular) component, accounting for the impact of unexpected factors on the time series [15].To discover meaningful rules for power assets, the data requires smoothing and counter-cyclical adjustment [16]. The processing steps are as follows:

Step 1: Trend-Cycle Estimation and Initial Smoothing

ISSN: 2664-9640

Perform trend-cycle analysis to obtain the product A of seasonal variation and random error $\hat{y}_{m6}$. Apply a centered moving average to the power grid asset data for smoothing:

$$\hat{y}_{m6} = \frac{(0.5y_{t-6} + y_{t-5} + y_{t-4} + \ldots y_{t+4} + y_{t+5} + 0.5_{yy+6})}{12} \tag{3}$$

For quarterly data, a two-period (e.g., two-quarter) moving average is used, expressed as:

$$\hat{y}_{q2} = \frac{(0.5y_{t-2} + y_{t-1} + y_t + y_{t+1} + 0.5y_{y+2})}{4} \tag{4}$$

The smoothed data $\frac{y}{\hat{y}}$ is then defined as:

$$\frac{y}{\hat{y}} = S \times e \tag{5}$$

where $\hat{y}$ is the normalized seasonal factor, S is the seasonal term, and since seasonality does not exist in the moving data, therefore $\hat{y} = 1$.

Step 2: Eliminate the error term and conduct quarterly analysis on the seasonal component S. The values corresponding to each quarter are called the normalized seasonal factors. After removing the long-term changes, the data A containing the seasonal and random error components is obtained. Normalize each seasonal factor to ensure that the mean of each quarter's indicator equals 1. That is:

$$zb_{im} = \frac{z_i \times 12}{\sum_{j=1}^{12} z_j} \tag{6}$$

Normalization processing of power grid asset data:

$$zb_{iq} = \frac{z_i \times 4}{\sum_{j=1}^{4} z_j} \tag{7}$$

Step 3: Remove the seasonal component from the raw data to obtain seasonally adjusted data.

Table 1. Cross-efficiency matrix

DMU	Target DMU				Average cross-efficiency
	1	2	⋯	n	
1	E_{11}	E_{12}	⋯	E_{1n}	$\bar{E}_1$
2	E_{21}	E_{22}	⋯	E_{2n}	$\bar{E}_2$
⋮	⋮	⋮	⋱	⋮	⋮
n	E_{n1}	E_{n2}	⋯	E_{nn}	$\bar{E}_n$

ISSN: 2664-9640

2.3. Risk Early Warning Model based on Association Rule Mining

The association rule mining algorithm was first proposed by Bezdek. Compared with other mining algorithms, the association rule mining algorithm is the most effective and computationally simplest algorithm. Firstly, a sample library of power grid asset system resources was established, and the risk warning was solved by using this method, thereby obtaining a new power grid asset system resource library. When the number of clusters C remains unchanged, it is found that the objective function of discovering association rules can be expressed as:

$$J(U,V,\Phi)=\sum_{q=1}^{c}\sum_{k=1}^{N}\mu_{q,k}^{W}\left\|\varphi(kT)-v_q\right\|^2 \tag{8}$$

In the formula, w is the exponential weight that affects the fuzzification of matrix $w\in(1,\ \infty)$, which is usually set to 2. $u_{i,k}$ needs to satisfy the following conditions:

$$\begin{cases}\sum_{q=1}^{c}\mu_{q,k}=1,\ \ k=1,2,\ldots,N\\ 0<\sum_{k=1}^{N}\mu_{q,k}<N,q=1,2,\ldots,c\end{cases} \tag{9}$$

Minimize the objective equation (8) to obtain equations (10) and (11) :

$$v_q=\frac{\sum_{k=1}^{N}\mu_{q,k}^{w}\varphi(kT)}{\sum_{k=1}^{N}\mu_{q,k}^{w}},\ \ q=1,2,\ldots,\ \ c \tag{10}$$

$$\mu_{q,k}=\frac{1}{\sum_{j=1}^{c}\left(\frac{\left\|\varphi(kT)-v_q\right\|}{\left\|\varphi(kT)-v\right\|}\right)^{2/(w-1)}},\ \ q=1,2,\ldots,c$$
$$k=1,2,\ldots,\ \ N \tag{11}$$

However, the analytical solutions of equations (10) and (11) cannot be effectively obtained. Therefore, the association rule mining algorithm is used to provide an iterative algorithm for approximately obtaining the minimum value of the objective function. On this basis, through iterative calculation, the membership degree matrix U and cluster center V of the power grid asset system were obtained. For c given number of clusters S_q, the parameter a to be identified can be determined through an algorithm:

$$s_q=[\frac{1}{p}\sum_{l=1}^{p}(c_q-c_l)^{T}(c_q-c_l)^{1/2}],\ \ q=1,2,\ldots,\ \ c \tag{12}$$

ISSN: 2664-9640

In the formula, p represents the number of nearest neighbors of the q -th cluster, and $c_l(l=1,2,\ldots,p)$ is the cluster center of each nearest neighbor of C_q. After preprocessing n time series data, the research was conducted in accordance with the time series technology described in Section 1.2. On this basis, a data mining method based on association rules is proposed. A new cycle is established based on the monthly mean value, and then the above method is used for association rule mining.

Suppose $X=\{X_1,X_2,\ldots,X_n\}$ is a new time series. A time window with a certain width of ω is applied to X, thereby obtaining a subsequence $Y_i=\{X_i,\ X_{i+1},\ldots,\ X_{i+\omega-1}\}$ with a length of X. The time window slides step by step from the beginning to the end of time series X, further forming subseries M_1, M_2,…, $M_{N-\omega-1}$ with a width of ω. Meanwhile, set $W(X,\omega)=\{M_i|i=1,2,\ldots,N-\omega+1\}$ as the subset of the sliding window of the time series X. $W(X,\omega)$ regards ω as $N-\omega+1$ point in x a-dimensional Euclidean space [17]. On this basis, after conducting a correlation analysis of the power grid asset indicators, some highly relevant power grid asset indicators were eliminated to simplify the model. The correlation coefficients of each asset indicator can be obtained by:

$$r_{x,y}=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-(\sum x_i)^2}\cdot\sqrt{n\sum y_i^2-(\sum y_i)^2}} \tag{13}$$

In the formula, x and y are two constant variables, and $r_{x,y}$ is the correlation coefficient of each variable.

3. Risk Early Warning Performance Test Results

3.1. Risk Early Warning and Prediction Changes

Simulation experiments were conducted using Matlab2020. The experimental equipment is a computer with an i7-7500 frequency of 2.90 GHz. The computer system is Windows10 64-bit Professional Edition, with 4G of memory, a main frequency of 3.8GHz, and a total of 1146 test samples. And conduct early warning and identification of common risks in power grid assets, namely, data leakage of power grid assets (700 test samples), abnormal financial data of assets (254 test samples), and data tampering of power grid assets (192 test samples).

The risk early warning and prediction results based on the association rule mining algorithm model are shown in Figure 1. Each row represents the actual category, and each column represents the risk warning category. The diagonal elements represent the number of correctly classified samples, and the non-diagonal elements represent the number of wrongly classified samples. The horizontal element in the last row and the vertical element in the last column respectively represent the early warning accuracy rate and recall rate of different risk categories of power grid assets. The last element represents the overall early warning accuracy rate of the model. The recall rates of data leakage, abnormal financial data and data tampering were 99.09%, 97.5% and 100.00% respectively, and the accuracy rates were 99.91%, 82.98% and 80.00% respectively. The overall classification accuracy rate was 99.04%. The accuracy rate of risk early warning for data leakage is relatively high. This is because data leakage has a significant impact on power grid assets and the number of data leakage samples is large. However, the accuracy rates of abnormal financial data and data tampering are relatively low, mainly due to the small amount of test data. It can be concluded from the confusion matrix that the risk early warning based on the association rule mining algorithm can effectively identify

ISSN: 2664-9640

risks such as data leakage, and the overall early warning accuracy rate is relatively high, which can meet the risk early warning requirements of power grid assets.

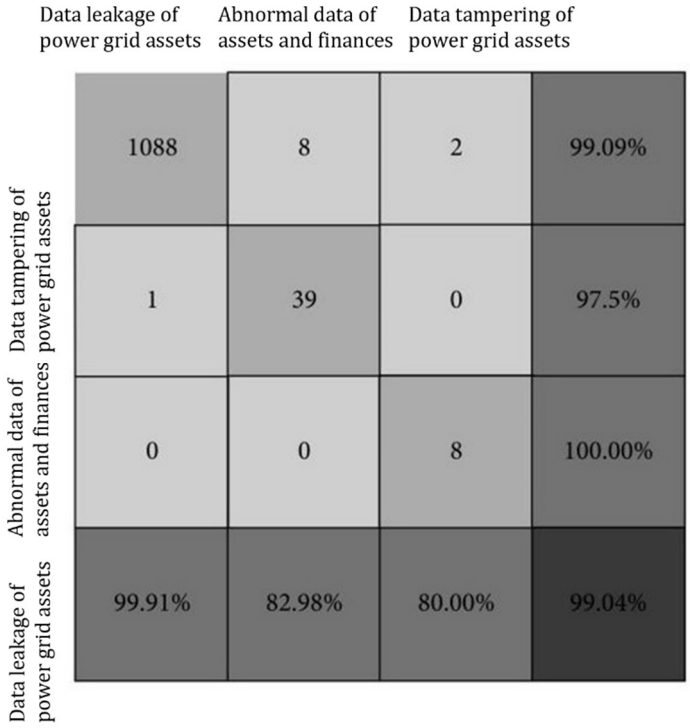

Figure 1. Confusion Matrix of Risk Warning Model

3.2. The Impact of Risk Warning Waiting Time on Warning Efficiency

It can be known from Section 3.1 that the recall rate of data tampering of power grid assets is 100%, while the early warning detection rate is only 80%. Therefore, it is necessary to further study the impact of association rule mining algorithms on the risk early warning performance in power grid assets.

When the risk early warning system detects the tampering of asset data, it will first conduct a secondary determination of the data. It will wait for a certain period of time to re-conduct risk early warning detection on the data. Moreover, during the asset data detection, the memory resources in the power grid asset system will be mobilized, thus having a certain impact on the detection rate capability of power grid asset risk early warning. It is necessary to further study the impact of waiting time on the efficiency of power grid asset risk early warning under different data sizes of power grid assets (100 kB, 400 kB, 600 kB).

It can be observed from Figure 2 that when the waiting time t increases from 10 ms to 200 ms, the efficiency of risk early warning detection decreases nonlinearly. The larger the data and the more the waiting time, the lower the efficiency of risk early warning detection. The main reason is that the larger the data, the greater the computational burden capacity of the system. Meanwhile, it can be observed that when the asset data is 100 kB, the detection efficiency is all greater than 95%. When the data size is 400 kB, the detection efficiency is all greater than 95%, while when the data is 600 kB, the detection efficiency is all greater than 92%. And the detection rates with a waiting time of 50ms were all greater than 97%. The larger the asset data and the longer the waiting time will have a certain impact on the detection efficiency. Therefore, it is suggested that the waiting time for power asset risk early warning be set within 50 ms, which can ensure the efficiency of data risk detection.

ISSN: 2664-9640

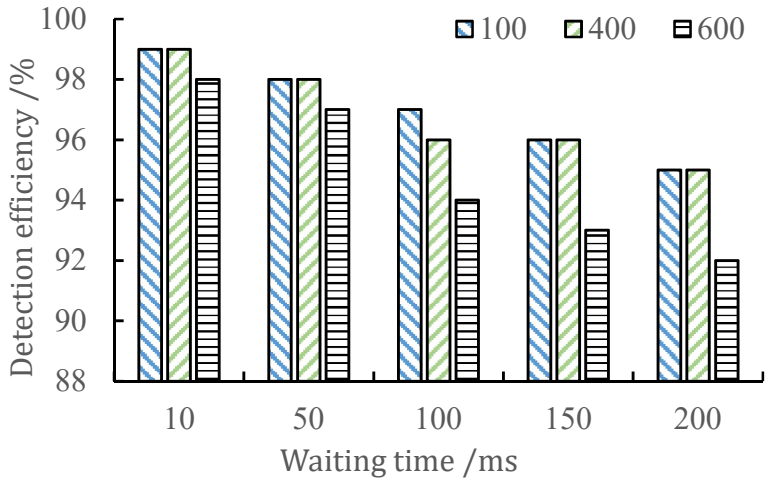

Figure 2. Effect of waiting time on the efficiency of risk detection

3.3. Analysis of Changes in Risk Warning Delay

The early warning delay effects of power grid assets of association rule mining algorithm, machine learning algorithm and extreme learning machine algorithm were compared on the CICIDS 2017 dataset. The delay results of three algorithms for risk early warning such as data leakage of power grid assets, abnormal financial data of assets and data tampering of power grid assets were further studied by using Matlab software.

Table 2. Results of Grid Asset Warning Delay Test for Three Systems (Unit: ms)

	Association rule mining algorithm			Machine learning algorithm			Extreme learning machine algorithm		
Iteration Frequency	Data leakage of power grid assets	Abnormal data of assets and finances	Data tampering of power grid assets	Data leakage of power grid assets	Abnormal data of assets and finances	Data tampering of power grid assets	Data leakage of power grid assets	Abnormal data of assets and finances	Data tampering of power grid assets
100	501	497	499	1235	1225	1287	1987	1954	1899
200	501	497	499	1209	1228	1288	1956	1985	1899
300	503	495	501	1225	1225	1289	1954	1954	1899
400	503	497	502	1228	1225	1288	1985	1985	1903
500	501	495	499	1225	1228	1289	1953	1954	1903
600	501	497	502	1228	1225	1288	1958	1985	1903
700	501	495	499	1225	1225	1289	1958	1954	1903
800	502	495	499	1228	1225	1289	1956	1985	1903

It can be known from Table 1 that there are significant differences in the early warning delay test results of the three algorithms for data leakage, abnormal data and data tampering. The number of iterations has a relatively small impact on the early warning delay results. With the increase of the number of iterations, the maximum variation range of the risk early warning delay time is only 10.56%. The average interval of early warning delay for data leakage, abnormal data and data tampering based on the association rule mining algorithm is 496 ms-502ms, and the minimum delay time is only 495 ms. The average early warning delays of machine learning algorithms and extreme learning machine algorithms for data leakage, abnormal financial data and data tampering both exceed 1200ms, which is twice that of association rule mining algorithms. Moreover, the average early warning delay times of the

ISSN: 2664-9640

three algorithms were 499 ms, 1246 ms, and 1943 ms respectively. The average delay time of the extreme learning machine algorithm was the longest, increasing by 74.32% compared with the association rule mining algorithm. Therefore, the risk early warning delay time of the proposed association rule mining algorithm is the shortest, indicating that the association rule mining algorithm can issue real-time early warnings for data leakage, abnormal data and data tampering, and achieve precise monitoring of power grid assets.

4. Conclusion

Power grid asset risk early warning is an important aspect of the intelligent transformation of the power system. Its core significance lies in achieving scientific and effective management of power assets through digital and intelligent means, and driving the intelligent upgrade of the power industry. The simulation study shows that the proposed association rule mining algorithm using time characteristics has a relatively high accuracy rate for risk early warning of data leakage, strong reliability, and reduces the risk of false alarms. It can provide a reference technical means for risk early warning in power grid assets and has certain application value in practice. Meanwhile, it is confirmed that the larger the power asset data and the longer the waiting time will have a certain impact on the efficiency of risk early warning detection. Therefore, it is recommended to set the waiting time within 50 ms. Compared with traditional machine learning algorithms and extreme learning machine algorithms, the association rule mining algorithm proposed in this paper has the shortest risk early warning delay time, indicating that the algorithm is both efficient and has a fast response speed. Thus, the time threshold for risk control can be moved forward, and the initiative can be gained in the "time battlefield", which means that decision-makers can start the emergency plan for handling the risks of power grid assets more quickly.

References

[1] Li Yan, Yang Lizhi, Zhang Lei, et al Research on Risk Early Warning of Power Grid Access Server Based on Time Series Database [J]. Electronic Design Engineering, 31(2023), 18, 123-126+131.

[2] Li Yanlin Intelligent Monitoring and Early Warning System for Risk Behaviors in Electromechanical Engineering Based on Power Grid Information Model [J] Journal of Shanghai Dianji University, 25(2022), 04, 234-238+248.

[3] Wang Jin, Pei Liang. Anomaly Detection and Multi-stage Risk Early Warning of Power Grid Regulation System Based on Deep Learning [J]. Journal of Shenyang University of Technology, 43(2021), 06, 601-607.

[4] CAI Guangcheng, Cao Lixia, Liu Xing, et al. Research on the Risk Early Warning System of Internal Audit in Power Grid Enterprises Driven by Knowledge Mining Technology [J]. China Internal Audit,(2021), 05, 31-36.

[5] Yang Jun, Wang Dong, Xu Jie, et al. Research on Risk Early Warning of Electric Power Production Safety Accidents Based on Data-Driven [J]. Electric Power Big Data, 22(2019), 04, 9-14.

[6] Chen Long, Hao Hanyong, Chao Yujian, et al. Research and Application of Intelligent Risk Early Warning Strategy for Electric Power Information Network Based on Big Data [J]. Electric Power Information and Communication Technology, 17(2019), 01, 18-24.

[7] Huang Wei, Huang Tingcheng, Wang Liyong, et al. Research Review on Typhoon Early Warning and Defense Framework of Power Grid Based on Situation Awareness [J]. Power System Protection and Control, 46(2018), 11, 162-169.

[8] Zhao Xiangyu, Wang Chengliang, Yao Gang, et al. Design and Development Application of Online Accident Risk Early Warning System for Power Grid Based on Accident Regulation of China Southern Power Grid [J] China Electric Power,49(2016), S1, 62-66.

ISSN: 2664-9640

[9] Chen Qiming, Chen Huayou. Research on the Early Warning Degree of Enterprise Financial Crisis Based on Analytic Hierarchy Process and Grey Fuzzy Evaluation Method [J] Industrial Technology & Economics, 30(2011), 03, 142-148.

[10] Guo Beibei, Ding Lijiang. Research on the Risk Distribution, Generation Mechanism and Countermeasures of Non-performing Assets in the Banking Industry: From the Perspective of Risk Early Warning [J] Journal of Changzhi University,39(2022), 01, 26-33.

[11] Du Tao, Wang Chaolong, Zhu Jing, et al. Monitoring and Anomaly Detection Technology of Transformer Equipment Operation Data Based on Clustering Algorithm [J]. Adhesive, 49(2022), 12, 137-140.

[12] Tang Qu, Bi Shengling. Data Integrity Analysis of Power Engineering based on Density Clustering Algorithm [J]. Bonding, 44(2020), 12, 74-77.

[13] Wang Dingfa Feature integration and Automatic Detection Technology of Abnormal Information Based on Cable Status Data [J]. Bonding,50(2023), 05, 188-192.

[14] Huang Wenjie, Qin Fangli, Deng Shuchi, et al. Research on the Intelligent Enhancement of Distribution Network Data Information Based on Load Balancing Aggregation Technology [J]. Adhesive,50(2023), 05, 193-196.

[15] Geng Jipu, Jiang Jinxia, Zheng Xiaoyan, et al. Big Data-driven Fault Risk Early Warning Method for Distribution Networks [J]. Electric Power Information and Communication Technology, 20(2022), 07, 41-49.

[16] Xie Jinghai, Jia Yike, Su Dongyu, et al. Real-time Early Warning Algorithm for Power Grid Operation Risks Considering Concurrent Data Streams [J]. Mechanical Design & Manufacturing Engineering, 51(2022), 05, 130-134.

[17] Xu Meiling, Liu Yijuan, Wu Xuexia, et al. Research and Application of Power Supply Service Risk Pre-Control System Based on Big Data [J]. Electric Power Big Data, 25(2022), 03, 66-73.