

Application of Improved Apriori Algorithm in Diagnosis of Abnormal Building Energy Consumption

Wang Yanwei^{1,2}, Zhang Hanyuan^{1,2,*}, Zhang Guiqing^{1,2}, Tian Feng^{1,2}, Ren Fei^{1,2}

¹School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, 250101, China

²Shandong Key Laboratory of Intelligent Buildings Technology, Jinan, 250101, China

Abstract: The analysis of building abnormal energy consumption is of great significance to the effective energy saving of buildings. To apply the relationship between the running status of building equipment and energy consumption to the diagnosis of abnormal energy consumption, an abnormal diagnosis method of building energy consumption based on the improved Apriori association rules is proposed. An improved Apriori algorithm is proposed for building energy consumption data with a large amount of data and multi-value attributes. The improved Apriori algorithm determines whether different attribute values of the same attribute data are in advance when generating candidate sets, reduces the number of comparisons, and improves the algorithm efficiency. By analyzing the abnormal energy consumption of the chiller in the refrigeration station of a commercial building, the superiority of the improved Apriori algorithm is proved, and the abnormal energy consumption is found, which verifies the feasibility and practicability of the proposed method.

1 Introduction

With the rapid progress and development of society, the problem of energy shortage and environmental deterioration is becoming more and more serious. Energy-saving and emission reduction has become a global consensus. Under the background of energy conservation and emission reduction, effective diagnosis of energy consumption data of public buildings, timely detection of unreasonable operating conditions, or abnormal energy consumption of building equipment, is conducive to optimize the building management scheme and control to reduce energy consumption.

With the establishment of more and more energy consumption monitoring platforms and the rapid development of data mining technology, energy consumption anomaly diagnosis is no longer limited to the framework of traditional expert diagnosis methods. It has developed from statistical threshold abnormal energy consumption judgment method to energy consumption anomaly diagnosis method based on data mining. Seem used GESD (Generalized Extreme Studentized Deviate) algorithm to detect abnormal energy consumption data based on the classification of energy consumption data by box plot [1]. Lin et al. considered the temperature factor and used simulation to judge the abnormality of energy consumption based on the deviation between the simulated energy consumption and the actual energy consumption [2]. Qing et al. identified unique energy consumption patterns through cluster analysis of historical energy consumption data, established a

decision tree of energy consumption patterns for the data, and judged energy consumption anomalies through outlier analysis of newly collected data and historical data of the same pattern [3].

However, the above method does not take into account the strong coupling between building equipment and the great influence of the operating state of equipment on energy consumption. Apriori algorithm has a good performance in mining the correlation between the equipment itself and between the equipment and the building subsystems from a large number of measured data and has been applied to the field of building energy conservation diagnosis. Cabrera et al. used the Apriori algorithm to successfully unearth the correlation between time, classroom occupancy, and wasted lighting energy consumption [4]. Shi et al. used Apriori association rule classification to diagnose chiller faults, analyzed the fault mechanism with the help of domain knowledge, and improved the reliability of the data-driven fault diagnosis method [5]. Li Guannan extracted the energy consumption rules of multi-online air conditioning with the Apriori association rule mining algorithm, explained the found energy consumption rules with the professional knowledge of refrigeration and air conditioning, and analyzed the reasons for their occurrence and the law of influence on energy consumption [6]. However, in practical application, the above research focuses on the analysis results and does not take into account the shortcomings of low efficiency and large resource consumption of the Apriori algorithm in mining large amounts of data.

To apply the correlation between equipment

* Corresponding author: Zhang Hanyuan; E-mail: zhanghanyuan18@sdjzu.edu.cn

operating parameters and energy consumption to the abnormal diagnosis of large-scale building operating data under different energy consumption modes, the traditional Apriori algorithm is improved based on the characteristics of building energy consumption data with multi-valued attributes. Whether the candidate itemset contains different attribute values of the same attribute can reduce the generation of unreasonable candidate item sets, save analysis time and improve algorithm efficiency. The improved Apriori algorithm is applied to different energy consumption modes to find the correlation between equipment operation parameters and energy consumption. The obtained rules are interpreted, and the normal association rule knowledge base is constructed. By finding the observation values that do not conform to the rules of the knowledge base, abnormal energy consumption can be detected, and the accuracy of abnormal energy consumption diagnosis can be improved. It has important significance for optimizing building operation management and building effective energy-saving significance.

2 Related knowledge introduction

2.1 Introduction of Apriori algorithm

The general form of association rules obtained by mining building energy consumption data is shown in $\{A \Rightarrow B, (sup, conf)\}$. A and B are item sets (sets of items) belonging to the building energy consumption data set, and $A \cap B = \emptyset$, sup (support) refers to the percentage containing $A \cup B$, and $conf$ (confidence) refers to the percentage containing A as well as B :

$$sup = support(A \Rightarrow B) = P(A \cup B) \quad (1)$$

$$confidence(A \Rightarrow B) = P(B | A) = P(A \cup B) / P(A) \quad (2)$$

Apriori algorithm is a classic algorithm for mining association rule sets of massive data, on which many improved and optimized algorithms are derived [7-8]. The basic idea is as follows: firstly, candidate itemsets are generated on the original dataset, and by comparing the minimum support set in advance, all non-empty subsets of frequent itemsets are also frequent itemsets, the candidate itemsets are deleted, and the iterative method of layer by layer search is used to find all

frequent itemsets in the dataset. Then, according to the minimum confidence, the corresponding association rules are derived in each frequent itemset.

The two main operations of the Apriori algorithm to get frequent itemsets include self-join and pruning. Self-join is to connect the set k -of frequent L_k with itself to generate the set $(k+1)$ - of the candidate C_{k+1} . Pruning is to delete the infrequent candidate set that does not meet the minimum support.

2.2 Data mining analysis of building energy consumption

Table 2 shows the sample data of the processed building energy consumption, which is derived from the freezing station of a large commercial center in Jinan. It is used to record the operation parameters of each equipment in the cold source system, including the temperature of chilled water supply and return water, instantaneous cooling capacity, instantaneous flow rate, condenser pressure, evaporator saturation temperature, current operation percentage, and the energy consumption of chillers and chilled water pumps. Finally, a total of 9 variables are utilized as listed in Table 1.

Since the Apriori algorithm can only accept type attribute variables, and the chiller operating data are all numerical variables, the numerical variables are discretized, that is, the value range of the attribute is divided into several intervals, which constitute the concept stratification of the attribute. As shown in Table 2, different letters indicate that the attribute is in a certain range. Take the chiller meter as an example: A2 means that the cooling meter is in the range of [67,97] kWh, and A1 means that the cooling meter is in the range of [35,66] kWh. It can be seen that the discrete building energy consumption data has multiple attribute values for each attribute, presenting typical multi-valued attribute characteristics. When mining building energy consumption data, each rule can contain a variety of attributes, but one attribute can only correspond to one value. Therefore, the candidate itemsets with the same attribute and different attribute values, such as $\{A1A2\}$ and $\{B1B2\}$, which appear in the process of generating frequent itemsets, are unreasonable. However, the comparison of these candidate sets will consume a lot of resources and increase the running time of the algorithm.

Table 1 Summary of the selected variables

Variable type	Abbreviation
Power consumption of chiller	P_{CH}
Power consumption of chilled water pump	P_{CHW}
Temperature of supply chilled water	T_{SCHW}
Temperature of return chilled water	T_{RCHW}
Instantaneous cooling power	ICP
Instantaneous flow	IFL
Pressure of condenser	PC
Saturation temperature of evaporator	$STEV$
Percentage of current running	PCR

Table 2 Examples of some building energy consumption data after discretization processing

<i>PCH</i>	<i>PCHW</i>	<i>TSCHW</i>	<i>TRCHW</i>	<i>ICP</i>	<i>IFL</i>	<i>PC</i>	<i>STEV</i>	<i>PCR</i>	<i>Mode</i>
A2	B2	C1	D1	E2	F2	G1	H1	I2	1
A2	B2	C1	D1	E2	F2	G1	H1	I1	1
A1	B2	C1	D1	E2	F2	G1	H1	I1	1
A1	B1	C2	D1	E1	F1	G3	H3	I1	1
A1	B1	C1	D1	E1	F1	G3	H3	I1	1

3 Improved algorithm introduced

3.1 Improved Apriori algorithm

According to the above analysis, the generation of candidate itemsets composed of the same attribute data is one of the reasons for the low efficiency of the Apriori algorithm. Because of, the idea of improving the Apriori algorithm is as follows:

When the candidate set k - ($k \geq 2$) is generated, the judgment condition of whether it is different attribute values of the same attribute data is added. Compare whether the same attribute data in two connectable frequent itemsets have different attribute values. If there are different attribute values, the candidate frequent itemset generated by the connection is unreasonable, and this connection will not be generated. On the contrary, this connection method can form a reasonable candidate itemset until all candidate frequent itemsets are found.

The improved Apriori algorithm can effectively delete the candidate itemsets generated by the connection between different attribute values of the same attribute, and reduce the number of comparisons when looking for frequent itemsets, thus improving the mining efficiency of the Apriori algorithm.

3.2 Improved Apriori algorithm example

Take the example data of building energy consumption in Table 2 as an example to show the process of generating 2- and 3- candidate sets by improved Apriori algorithm:

Assuming that the minimum confidence is 0.3, the frequent 1- itemset has {A1}, {A2}, {B1}, {B2}, {C1}, {D1}, {E1}, {E2}, {F1}, {F2}, {G1}, {G3}, {H1}, {H3}, {I1}. A total of 105 candidate itemsets were obtained by L_1 self-connection, including {A1A2}, {B1B2}, {E1E2}, {F1F2}, {G1G3}, {H1H3} unreasonable candidate itemsets composed of the same attribute data with different attribute values. These 6 candidate itemsets were deleted. The remaining 99 candidate item sets were pruned to delete the candidate itemsets that did not meet the minimum support, and 63 frequent item sets were obtained.

A total of 157 candidate itemsets were generated by self-join, including 4 unreasonable candidate itemset {D1E1E2}, {D1F1F2}, {D1G1G3}, {D1H1H3}, which were deleted. Then pruning is performed to get frequent

3- itemsets.

Through the above process of generating frequent 2- and 3- sets, it can be seen that the improved Apriori algorithm can delete candidate itemsets generated by the connection of different attribute values of the same attribute, reducing the number of comparisons, especially when the amount of data is large. In the case of a large number of attribute values, at this time, the candidate itemset generated by the same attribute data connection is greatly increased, and the improvement effect of the algorithm is more significant.

4 Introduction of abnormal building energy consumption diagnosis model

The building abnormal energy consumption diagnosis model based on the improved Apriori algorithm effectively integrates the energy consumption data and operation parameters of different equipment, excavates the association between equipment and between equipment operation parameters and energy consumption, selects and evaluates meaningful strong association rules to establish the normal operation association rule base, and judges whether the observed value is abnormal energy consumption according to the rules of the rule base. Mainly includes three stages: data preprocessing, improved Apriori algorithm mining, the establishment of normal rule base and diagnostic applications.

Data preprocessing mainly includes processing the obvious abnormal value, missing value, and inconsistent data in the building energy consumption data, and normalizing the data. The purpose is to improve the quality of data mining results so that data mining results can guide practice.

Then, by improving the Apriori association rule algorithm, the minimum support and minimum confidence are set to mine the hidden knowledge in the processed data of different building operation modes, and the corresponding association rules are derived.

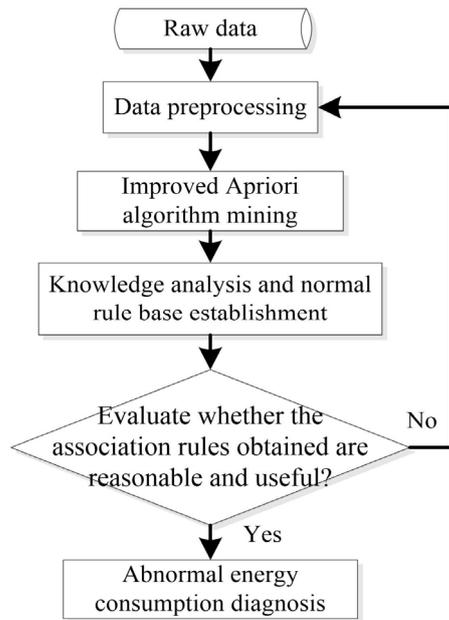


Fig. 1 Framework process of building abnormal energy consumption diagnosis model

Finally, the professional knowledge is applied to analyze and evaluate the obtained association rules, and the reasonable and useful association rules are established into the normal rule knowledge base. By checking the consistency between the new observation value and the normal rule knowledge base, whether the building energy consumption is abnormal is judged. Fig. 1 shows the framework flow chart of the abnormal building energy consumption diagnosis model.

5 Experimental results and analysis

The experimental data are 4032 pieces of data from the freezing station of a large commercial center in Jinan. After data preprocessing, two different energy consumption operation modes are obtained by using a clustering algorithm.

To improve the efficiency of association rules mining and the reliability of the knowledge found, the improved

Apriori algorithm is used to mine the association rules in two modes.

5.1 Improved Apriori association rule mining results

To verify the superiority of the improved Apriori algorithm, the improved algorithm is compared with the traditional Apriori algorithm, and the data is mined under different support degrees. The time-consuming of the two algorithms is shown in Fig. 2.

As can be seen from Fig. 2, the effect of the improved Apriori algorithm is better than that of the traditional algorithm, especially in the case of the smaller support, the more irrational candidate sets caused by the same attribute data connection, which affects the efficiency of the algorithm.

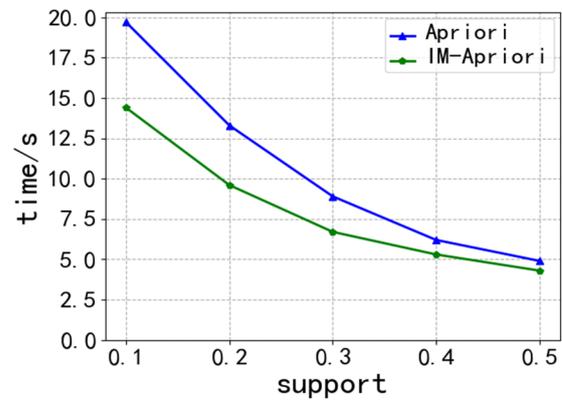


Fig. 2 Comparison of time consumption of the two algorithms under different support degrees

The improved Apriori algorithm is used to mine the experimental data with the minimum support of 0.1 and the minimum confidence of 0.9. Part of the results are shown in Table 3. The greater the support in Table 3, the more frequent it appears. The closer the confidence is to 1, the higher the reliability is.

Table 3 Shows part of the association rules

No.	Antecedent	Consequent	Support	Confidence	Mode
1	$T_{SCHW}=[9.69,10.95],S_{TEV}=[6.4,7.6]$	$P_{CH}=[67,97]$	0.372	0.96	1
2	$T_{RCHW}=[16.01,18.77]$	$P_{CH}=[67,97]$	0.281	1	1
3	$T_{SCHW}=[9.69,10.95],P_C=[815.9,909.6]$	$P_{CH}=[67,97]$	0.385	0.96	1
4	$T_{SCHW}=[7.89,8.62],I_{FL}=[144.3,199.3]$	$P_{CH}=[35,66]$	0.346	0.95	1
5	$T_{SCHW}=[13.4,15.23],P_C=[718.6,814.4]$	$P_{CH}=[35,66]$	0.382	0.95	1
6	$T_{SCHW}=[6.76,8.35],I_{CP}=[516.1,756.9]$	$P_{CH}=[30,46]$	0.358	0.94	2
7	$T_{RCHW}=[10.9,13.2],I_{FL}=[195.8,230.1]$	$P_{CH}=[47,65]$	0.337	0.95	2
8	$I_{FL}=[7.89,8.62],P_{CR}=[51,74]$	$P_{CHW}=[5.8,12.7]$	0.328	0.96	1

9	$P_{CH}=[35,66], T_{RCHW}=[14.7,15.9]$	$P_{CHW}=[1.3,5.7]$	0.376	0.97	1
10	$P_{CH}=[30,46]$	$P_{CHW}=[0.4,3.8]$	0.305	0.97	2

5.2 Rule analysis and diagnostic application

From the obtained rules, T_{SCHW} , T_{RCHW} , S_{TEV} , P_C , I_{FL} , I_{CP} and P_{CR} can reflect the correlation with chiller energy consumption.

From the basic principle of chiller refrigeration: the low-temperature refrigerant exchanges heat with the chilled water from the end of the user in the evaporator, absorbs the heat of the chilled water to cool the chilled water, and then transports the chilled water to the end of the user, and uses the low-temperature chilled water to absorb the heat in the room. From the point of view of chilled water supply and return water temperature, the saturated temperature and pressure of the evaporator are directly related to the heat brought by chilled water into the evaporator. When the refrigeration demand is large, the return water temperature of the chilled water in the evaporator increases, which makes the saturated temperature of the evaporator increase and the corresponding evaporator pressure increase. In turn, when the refrigeration demand is small, the return water temperature of chilled water decreases, and the evaporator saturation temperature and evaporation pressure also decrease. The saturation temperature and evaporation pressure of the generator will also decrease. The evaporation saturation temperature and pressure have a great impact on the evaporation efficiency, and the refrigerant evaporation efficiency in the evaporator has a direct impact on the operation of the chiller, which ultimately reflects the impact on the energy consumption of the chiller. Similarly, the saturation temperature and pressure of the condenser determine the condensing efficiency, and the condensing efficiency of the refrigerant in the condenser also has a direct impact on the operation of the chiller, so the correlation between the condenser pressure and the energy consumption of the chiller is established.

The three parameters of I_{FL} , I_{CP} , and P_{CR} all reflect the change of load rate. Load rate is an important factor

affecting the operation of chillers. The results show that the flow rate of chilled water, the energy consumption of the chilled water pump, and the energy consumption of the chiller are influenced by each other. From the mining rules, the influence of the flow rate of chilled water, the running state of the chiller, and the energy consumption of the chiller on the energy consumption of the chilled water pump also confirms that there is a complex coupling relationship between them. The above rules can be explained from domain knowledge, which is reasonable and effective, and a normal rule base can be established.

The normal rule base is used to monitor the operation data of the chiller, and some problems in the operation of the cold source system are found. First, the meter is faulty. During the above data checking process, the energy consumption of the chiller and the refrigerating pump from 11:30 to 19:45 on September 27 is 0, but the operating parameters of the chiller are all within the range of each rule of Mode 2, that is, the chiller is Normal operation status, so it can be judged that the meter is malfunctioning during this time. Second, the abnormal energy consumption of the chiller. The observations in operation mode 1 from 16:00 to 17:15 on August 23 violated Rule 5 of Table 3. As shown in Table 4, during this time, the return water temperature fluctuates in the range of [13.94,14.28]°C, the current percentage range is between [66,73], but the energy consumption of the chiller is between [75,83], Inconsistent with the above rules. Observing the data, it can be found that compared with the normal data, the condenser pressure is in the range of [718.6,814.4] Kpa, and the condenser pressure rises under abnormal energy consumption and is in the range of [891.3,904.4] Kpa, and the condenser saturation temperature rises at the same time. In this case, the compression function of the compressor increases, the compression ratio becomes larger, and the refrigeration capacity is relatively reduced, which increases the energy consumption of the chiller

Table 4 Display of abnormal data on August 23

Time	P_{CH}	P_{CHW}	T_{SCHW}	T_{RCHW}	I_{CP}	I_{FL}	P_C	S_{TEV}	P_{CR}	Mode
16:00	76	7.1	9.12	14.1	850.3	171.4	891.3	6.1	66	1
16:15	78	6.8	9.05	13.94	866.2	175.3	901.2	6.2	71	1
16:30	75	6.9	9.15	14.14	849.2	180.5	904.4	6.1	72	1
16:45	79	6.8	9.23	14.21	861.7	182.4	894.8	6.1	68	1
17:00	83	7.1	9.18	14.16	854.6	184.6	900.2	6.2	73	1
17:15	79	7.1	9.19	14.28	842.8	170.8	903.9	6.2	70	1

6 CONCLUSION

In this paper, based on the characteristics of a large amount of building energy consumption data and

multi-valued attributes, the improved Apriori algorithm is used to find the correlation between equipment operation parameters and apply it to large-scale building energy consumption data anomaly diagnosis. To solve the problem of high time complexity of Apriori

algorithm caused by unreasonable candidate set generated by the same attribute data connection, this paper uses the method of judging whether it is the same attribute data connection to reduce the generation of the unreasonable candidate set, to improve the efficiency of the algorithm. The experiment and test on the chiller operation data of the refrigeration station verify the effectiveness of the improved Apriori algorithm, and the abnormal energy consumption is successfully detected in the monitoring of the chiller operation data, which proves the feasibility of the method.

Acknowledgments

This research was partly funded by the Shandong Provincial Major Science and Technology Innovation Project (2019JZZY010115), the National Natural Science Foundation of China (62003191), the Natural Science Foundation of Shandong Province (ZR2020QF072), and the Doctoral Research Foundation of Shandong Jianzhu University (XNBS1821).

References

1. Seem, J.E. (2007) Using intelligent data analysis to detect abnormal energy consumption in buildings. *J. Energy. & Buildings*, 39(1): 52-58.
2. Lin, G., Claridge, D.E. (2015) A temperature-based approach to detect abnormal building energy consumption. *J. Energy. & Buildings*, 93: 110-118.
3. Qing, X.X., Xiao, D., Wang, B. (2012) Data mining method for real-time monitoring of energy consumption. *J. Journal. Of. Chongqing. University*, 35(07): 133-137.
4. Cabrera, D.F.M., Zareipour, H. (2013) Data association mining for identifying lighting energy waste patterns in educational institutes. *J. Energy. & Buildings*, 62(Jul.): 210-216.
5. Shi, D.L., Liu, J.Y., Li, K.N., et al. (2020) Fault diagnosis of chiller based on association rule classification. *J. Journal. Of. Chiller. Engineering*, 41(5): 66-75.
6. Li, G.N., Hu, Y.P., Chen, H.X., et al. (2017) Data partitioning and association mining for identifying VRF energy consumption patterns under various part loads and refrigerant charge conditions. *J. Applied. Energy*, 39(1): 846-861.
7. Wan, X., Hu, N.S., Han, P.F., et al. (2016) Research on the application of big data mining technology to the operation performance optimization of steam turbine. *J. Chinese. Journal. Of. Electrical. Engineering*, 36(2): 459-467603.
8. Chen, Y., Li, S.N., Zhang, L., et al. (2019) Defect correlation analysis of secondary equipment in Smart Substation Based on improved Apriori algorithm. *J. Protection. And. Control. Of. Power. System*, 47(20): 135-141.