Mining Research on Vibration Signal Association Rules of Quayside Container Crane Hoisting Motor Based on Apriori Algorithm

Chencheng Yang, Gang Tang and Xiong Hu
Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China
yclencheng@126.com

Abstract. Shore-hoisting motor in the daily work will produce a large number of vibration signal data, in order to analyze the correlation among the data and discover the fault and potential safety hazard of the motor, the data are discretized first, and then Apriori algorithm are used to mine the strong association rules among the data. The results show that the relationship between day 1 and day 16 is the most closely related, which can guide the staff to analyze the work of these two days of motor to find and solve the problem of fault and safety.

1. Introduction
In order to adapt to more and more ships, shorten loading and unloading time and improve efficiency, the quay crane, which plays an important role in the port operation, is developing towards high-speed operation and heavy lifting capacity, so the hoisting motor is easy to produce the problem of large load, big impact and big sway [1]. Hoisting motor vibration signal data is very valuable and analytical value, but these data are often disorganized and there is no obvious feature, so the need for a hidden relationship between data mining methods Apriori algorithm is a frequent itemset mining algorithm for mining association rules, which uses join and pruning methods to generate a small number of candidate sets and then generate frequent itemsets. Finally, strong association rules are generated, which are widely used in various fields of social life the Apriori algorithm can mine the association rule in the motor vibration signal data, analyze the mechanical state of the motor, and find out the motor fault and safety hazard [2].

2. The principle of Apriori algorithm
There are two events A and B. The support of association rule $A \Rightarrow B$ refers to the proportion of all transactions in the database that contain the event AB, that is the probability $P(A \cup B)$. The confidence of association rule AB is the number of transactions that contain both AB and include A ratio of the number of transactions, that is the probability $P(B|A)$. Apriori algorithm is a priority algorithm. Every time the database is scanned, it calculates all the support in the database. The items that support less than the minimum value are set to form a frequent itemset L1. Then it scans the database and generates a frequent itemset L2. The kth scan generates a frequent itemset
Lk of length k and the (k+1)th scan generates a frequent itemset Lk+1 of length k+1. The items in the candidate frequent itemset Ck+1 are reduced according to the pruning properties of the algorithm and scan the database to generate frequent itemsets according to the items in Ck+1 [3-4]. Apriori algorithm is an iterative way. First, it draws C1 from each item in the database. Then it draws the frequent itemset L1 from the minimum support [5-6]. Then it finds the combination Ck+1 of Lk and L1. Next it finds Lk+1 according to the minimum support and repeats this cycle until k-order. Finally, the confidence of itemset Lk is calculated. If the confidence is greater than the minimum confidence, it is a qualified rule. Method flow chart is shown in Figure 1 [7].

Figure 1. Apriori algorithm data mining schematic.
3. Data association rule mining process

Data association rules mining process includes preparing data, discretizing data, reading Matlab and running programs, generating frequent itemsets, generating strong association rules, analyzing strong association rules and so on.

The vibration signal data of 28 days were selected from 2009.12.28-2010.01.03 and 2010.01.18-2010.02.07, and these data were used as the original data of association rule mining. Because of the large amount of data, in order to facilitate the mining association rules and procedures for the operation, all data is divided into 28 columns (6 bits, respectively, hours, minutes and seconds), part of the data shown in Table 1:

| time/day | 09.12.28 | 09.12.29 | 09.12.30 | 09.12.31 | 10.01.01 | 10.01.02 |
|----------|----------|----------|----------|----------|----------|----------|
| 000001   | 1.348226717 | 1.461543966 | 0.147767787 | 1.732512912 | 1.988745204 | 0.215233879 |
| 000021   | 0.557416152 | 1.111776506 | 0.077004680 | 1.456330582 | 2.260567322 | 0.183927996 |
| 000031   | 1.051686803 | 1.776346064 | 0.079273979 | 0.809488764 | 1.830401507 | 0.174495186 |
| 000041   | 1.641297123 | 1.450069932 | 0.077004680 | 1.545590266 | 1.296293009 | 0.213564406 |
| 000051   | 0.978158806 | 1.242783452 | 0.134116793 | 1.456330582 | 2.608663835 | 0.141376765 |
| 000101   | 2.488720126 | 1.292546795 | 0.134116793 | 0.967015838 | 2.608663835 | 0.141376765 |
| 000121   | 1.4078731   | 1.850598386 | 0.106696628 | 0.610512535 | 0.549463057 | 0.273129161 |
| 000131   | 2.304847972 | 1.144550309 | 0.114081028 | 1.45902719  | 0.561034852 | 0.236489105 |
| 000141   | 1.671902493 | 1.253938726 | 0.073675301 | 1.45902719  | 6.87629457  | 0.251650358 |
| 000151   | 6.599175619 | 0.807598386 | 0.115374575 | 1.465584009 | 1.120279761 | 0.106547819 |

Because the association rules analysis algorithm discretizes the data, all transactions need to be pretreated, and the data is discretized by the definition of association rules. The commonly used methods are Boolean discretization and multi-valued discretization [8-9]. Boolean discretization, the data in the normal range of data is mapped to 0, in a non-normal range of the mapping is 1 (that is, the emergence of early warning state is 1, otherwise 0), where the data is greater than or equal to 1, the data of less than 1 becomes 0. The discretization result is shown in Table 2:

| time/day | 09.12.28 | 09.12.29 | 09.12.30 | 09.12.31 | 10.01.01 | 10.01.02 |
|----------|----------|----------|----------|----------|----------|----------|
| 000001   | 1        | 1        | 0        | 1        | 1        | 0        |
| 000021   | 0        | 1        | 0        | 1        | 1        | 0        |
| 000031   | 1        | 1        | 0        | 0        | 1        | 0        |
| 000041   | 1        | 1        | 0        | 1        | 1        | 0        |
| 000051   | 0        | 1        | 0        | 1        | 1        | 0        |
| 000101   | 1        | 1        | 0        | 0        | 1        | 0        |
| 000121   | 1        | 1        | 0        | 0        | 0        | 0        |
| 000131   | 1        | 1        | 0        | 1        | 0        | 0        |
| 000141   | 1        | 1        | 0        | 1        | 1        | 0        |
| 000151   | 1        | 0        | 0        | 1        | 1        | 0        |

Put the data of Boolean discretization into Matlab and save it as a file that can be transferred directly. Finally, the association rule mining based on Apriori algorithm can calculate the degree of support and
confidence between different time points of each day, and get the frequent itemsets and strong association rules[10-11]. Set a support and confidence threshold min_sup = 0.35, min_conf = 0.6. The itemsets whose support degree is greater than the threshold are frequent itemsets, and the strong itemsets whose confidence is greater than the threshold are strong association rules. Through Matlab calculating support The frequent itemsets are shown in Table 3.

| Number | Frequent itemsets | Support  |
|--------|-------------------|----------|
| 4      | 1 13              | 0.3500   |
| 6      | 1 16              | 0.4164   |
| 12     | 13 16             | 0.3504   |

And then calculate the confidence of each frequent itemset and compare it with the minimum confidence threshold. The result shows that the frequent itemsets of number 7 do not meet the requirements, so 13 strong association rules are finally obtained.

From the strong association rules, we can see that the probability of early warning signals of two-day lifting motors with strong association rules is higher at the same time, and there is a strong correlation between them, among them, the first and the 16th day (2009.12.28 and 2010.01.26) are most closely related.

The above results show that the two days the motor is in a similar working environment and working conditions, the two days may have a bad weather conditions or the motor is in high-load working condition, which can re-analysis of these two days of motor work, Draw conclusions and guide the relevant personnel in a timely manner to resolve failures and security risks.

### 4. Conclusion

In order to verify the results, using Matlab to draw the 1st day and 16th day of the data fitting diagram, as shown in Figure 2:

![Data fitting diagram](image)

From the figure, it can be seen that the distribution of the data in these two days is similar, which proves the correctness of the mining association rules.
The vibration signal data of the hoisting motor of the quayside crane is an important reference for analyzing the mechanical state of the motor, and the association of the vibration signal data with the association rules is an important method to evaluate the working status of the motor, and to prevent and reduce the fault. The association rules between the mechanical data of the motor provide a direction for processing the huge data of the motor, and can monitor and evaluate the working state of the motor. It can also find out the potential faults and safety hazards ahead of time, and has important reference to the relevant staff significance, in later studies can be more than a number of properties of the measurement points to analyze the data to be more accurate and comprehensive association rules.

Acknowledgements
This work was supported by the National Natural Science Foundation of China (No. 31300783), China Postdoctoral Science Foundation (No. 2014M561458), Doctoral Fund of the Ministry of Education Jointly Funded Project (No. 20123121120004), the Shanghai Maritime University Research Project (No. 20130474), the Shanghai Top Academic Discipline Project- management science & engineering, and the high-tech research and development program of China (No. 2013A2041106).

References
[1] Wang Zhi-xin. Data Mining and Condition Recognition on Characteristic Data of Quayside Container Crane[D]. Shanghai: Shanghai Jiao Tong University, 2008:15-23.
[2] Xiaoyan Cui, Shimeng Yang, Darning Wang. An Algorithm of Apriori Based on Medical Big Data and Cloud Computing[C]. 2016 4th International Conference on Cloud Computing and Intelligence Systems. IEEE Conference Publications, 2016: 361-365.
[3] Liu Hua-ting, Guo Ren-xiang, Jiang Hao. Research and Improvement of Apriori Algorithm for Mining Association Rules[J]. Computer Applications and Software, 2009, 26(1) : 146-149.
[4] Cui Guan-xun, Li Liang, Wang Ke-ke, et al, Research and Improvement of Apriori Algorithm for Mining Association Rules[J]. Computer Applications, 2010, 30(11) : 2952-2955.
[5] Huang Chang-hai, Gao De-yi, Hu shen-ping, et al, Association Rule Analysis of Vessel Traffic Accidents Based on Apriori Algorithm[J]. Journal of Shanghai Maritime University, 2014, 35(3) : 19-22.
[6] Wang Yu, Zhang Wei-hong, Liu Yu. Application of Association Rules in Information Accessibility Website Based on Apriori Algorithm[J]. Journal of Jilin University, 2013, 31(1) : 101-106.
[7] Jon Hills, Anthony Bagnall, Beatriz Iglesia, et al, BruteSuppression: A Size Reduction Method for Apriori Rule Sets[J]. Journal of Intelligent Information Systems, 2013, Vol.40 (3), pp.431-454.
[8] Lamine M. Aouad, Nhien-An Le-Khac, Tahar M. Kechadi. Performance Study of Distributed Apriori-like Frequent Itemsets Mining[J]. Knowledge and Information Systems, 2010, Vol.23 (1), pp.55-72.
[9] Dong Juan Gu, Lei Xia. A Novel and Improved Apriori Algorithm[J]. Applied Mechanics and Materials, 2015, 543-546.
[10] Guo Min Zhuang. Simulation Environment of Virtual Visualization Technology Based on Apriori Algorithm[J]. Applied Mechanics and Materials, 2014, 1131-1135.
[11] Niu Zhendong, Nie Yaoxin, Zhou Qian, et al, A Brain-region-based Meta-analysis Method Utilizing the Apriori Algorithm[J]. BMC neuroscience, 2016, 23.