The Principle of Fuzzy Negative Association Rule and Its Experiment Simulation

Zhaopeng Zhang
Datang Northeast Electric Power Test & Research Institute, Changchun, Jilin, China
gongyuanyang@neepu.edu.cn
*Corresponding author’s e-mail: 1004084545@qq.com

Abstract. Energy conservation optimization control is an important way to reduce the energy emission and save energy in coal-fired power plants. There is a large amount of potential knowledge and information in the massive data in the real-time historical database of thermal power plant, which can be used for the control of energy conservation and optimization. This paper is mainly around the fuzzy association rules based on the traditional method in power plant, the application of energy saving optimization control for fuzzy association rules in specific research, and the simulation test data of the power plant operation.

1. Introduction
An important means of thermal power production to improve its economy is to establish accurate mathematical model of the system through the optimization control unit operation and the traditional plant operation parameter optimization method, but the application is very limited. Data mining technology as a new discipline, can not only solve this problem effectively under the condition of little human intervention, but also can process complex data information and extract the knowledge about the rules. The rule of fuzzy association rules is a widely used method in data mining.

Negative association rules is actually a complement of association rules as well as the traditional association rule data mining algorithm, and some parts of invalid or meaningless association rules can be eliminated through improvement, but there is a lot of limitations and negative association rules can make use of the contradiction between the positive and negative association rules can eliminate is association rules on the other hand, can also generate new negative association rules, which plays a key role for association rules mining.

2. The basic definition of negative association rule
Define I={I1,I2,I3,I4….In}as a collection or set to projects set, in which IK(1<k<n) is a separate project, transaction X, Y loophole is not set belong to I, a type of association rule, such as implication, is a negative association rule. In a negative association rule, three conditions must be met for the negative association rule to be meaningful.The three conditions mentioned above can be listed as follows:

Firstly, the intersection of X and Y is not null. If the intersection of X and Y is empty, then X and Y are two independent transactions, and the two do not interact with each other. Secondly, the support degree of X and Y should be greater than the minimum support degree, which is to ensure that the negative association rule makes sense in the probability statistics; finally, the sum of X and -Y (or -X and Y) is greater than minimum support.
Theorem 1: set A and B are two subsets of item \( I \), and the intersection of A and B is not null:

- \( \text{support}(-A) = 1 - \text{support}(A) \)
- \( \text{support}(A \cap -B) = \text{support}(A) - \text{support}(A \cap B) \)
- \( \text{support}(-A \cap B) = 1 - \text{support}(A) - \text{support}(B) + \text{support}(A \cap B) \)

Theorem 2: set A and B are two subsets of the set \( I \), and the intersection of A and B is not null:

- \( \text{correlate}(A \Rightarrow -B) = 1 - \text{correlate}(A \Rightarrow B) \)
- \( \text{correlate}(-A \Rightarrow B) = (\text{support}(B) - \text{support}(A \cap B)) / (1 - \text{support}(A)) \)
- \( \text{correlate}(-A \Rightarrow -B) = 1 - \text{correlate}(-A \Rightarrow B) \)

According to the above two theorems, the support and confidence of the negative association rules can be calculated, which plays a key role in the mining algorithm.

3. A negative association rule algorithm based on apriori algorithm

In this paper, we present the pseudo-code of negative association rules, and introduce some concepts to make the mining of negative association rules more intuitive, it should be pointed out that, this algorithm is not only about the mining negative association rules, but also expand to include the negative association rules, the association rules algorithm is as follows:

Input: frequent itemset \( I \), minimum confidence MC (min confidence), minimum support ms (min support);

Output: association rule set \( R \):

for(itemset A and B \( \in I \))

\( \{ \)

\( \text{Int} \ X = P(A \cup B) / P(A)P(B) \)

If(confidence \( (A \Rightarrow B) > \text{mc} \))

\( \{ \)

\( R = R \cup \{A \Rightarrow B\} \)

\( \} \)

If(confidence \( (-A \Rightarrow -B) > \text{mc} \))

\( \{ \)

\( R = R \cup \{A \Rightarrow -B\} \)

\( \} \)

If \( (X < 1) \)

\( \{ \)

If(confidence \( (-A \Rightarrow B) > \text{mc} \))

\( \{ \)

\( R = R \cup \{-A \Rightarrow B\} \)

\( \} \)

If(confidence \( (A \Rightarrow -B) > \text{mc} \))

\( \{ \)

\( R = R \cup \{A \Rightarrow -B\} \)

\( \} \)

Return \( R \)

The above algorithm can use the correlation judgment to generate the relevant rules, so as to avoid the conflicting conclusions. At the same time, this algorithm only adds some judgment control statements to the original association rules mining algorithm, so the algorithm itself is not complicated.
4. Algorithm verification experiment

Based on the real-time data of a 325mw subcritical reheat condensing steam turbine in a power plant in Shanghai, under the condition of the best operating range, the working conditions of some important regulating parameters in the operation of the boiler are mined, in order to be able to clearly show the mining process, the example of single parameter’s oxygen drainage mining is selected.

The load of boiler is divided into 10 load intervals according to the natural condition, and the result of the division is:

\[[141.25,154.85],[154.94,169.82],[169.85,186.7],[186.74,203.5],[203.59,221.52],[221.55,240.98],[241.21,260.83],[260.91,279.12],[279.2,296.47],[296.54,320.5]\].

The coal quality coefficient, which reflects coal quality, is divided into three ranges: excellent, middle and poor, corresponding to the numerical interval:

\[[1.0367,1.3186],[1.3197,1.7083],[1.715,2.20]\].

Because only one week of data mining is needed, the change of environmental temperature parameter range is small, through the data of temperature maximum and minimum value between 15.92 and 24.8 and the temperature changes of 10℃within a week, so, there is no need to divide the ambient temperature. In order to facilitate the presentation of mining, we define condition 1 loads \[[154.94, 169.82], [15.92, 24.8]\] temperature, coal quality coefficient of the week running data records show 2314 jobs under the condition of the working of condition 1.

The starting data of the correlation rules is collected. The starting data of the mining of the association rules is as shown in table 1.

| External working condition parameter | State parameter |
|--------------------------------------|-----------------|
| Load | Ambient temperature | Coal quality coefficient | Flue gas oxygen content | Efficient |
| 169.36 | 19.28 | 1.49 | 6.97 | 92.06 |
| 168.91 | 18.89 | 1.54 | 7.41 | 91.71 |
| 169.19 | 18.93 | 1.54 | 7.48 | 91.68 |
| 167.72 | 18.99 | 1.54 | 7.55 | 91.65 |
| 168.25 | 18.96 | 1.52 | 7.62 | 91.62 |
| 166.98 | 18.92 | 1.50 | 7.67 | 91.59 |
| 166.26 | 18.92 | 1.50 | 7.67 | 91.58 |
| …… | …… | …… | …… | …… |
| 160.80 | 19.29 | 1.48 | 6.33 | 91.71 |

According to the procedure in the third chapter, the FCM algorithm is used to divide the smoke oxygen quantity of the boiler into low oxygen content, middle and high.

| Low oxygen uptake | The lower threshold range | The higher threshold range | Records | Rate |
|------------------|--------------------------|---------------------------|---------|------|
| 5.97             | 6.66                     | 355                       | 0.1534  |
Middle oxygen uptake 6.67 7.05 1183 0.5112
High oxygen uptake 7.06 8.11 776 0.3354

In the same way, the working condition 2 is excavated, and the final result is as follows:

![Figure 1. The result of mining](image)

According to the simulation results shown in table 3, we can draw the conclusion that, after introducing concept of correlation of negative association rules, the number of association rules mined by the algorithm is obviously smaller than that of the Apriori algorithm, which filters out the correlation of rules that do not conform to the new judgment system, and improves the accuracy of rule mining, the negative association rules Mining Algorithm is introduced to mine the negative association rules, which indicates that the correlation Judgment Algorithm is effective.

| Algorithm         | Apriorialgorithm | Algorithm of Negative association rule |
|-------------------|------------------|----------------------------------------|
| Number of positive association rule | 25               | 18                                     |
| Number of negative association rule | Null             | 34                                     |

5. Conclusions
This paper mainly discusses the origin of the negative association rules and the definition of negative association rules. In this paper, the correlation based negative association rule algorithm is discussed. The validity of the Algorithm is demonstrated theoretically and practically, which provides a theoretical basis for the further study of negative association rules.
References

[1] Johannes, G., Raghu, R., Venkatesh, G. (2000) Rainforest-A framework for fast decision tree construction of large datasets. Data Mining and Knowledge Discovery, 4(3):127-162.

[2] Malek, M., Kadima, H. (2013) Searching frequent itemsets by clustering data: towards a parallel approach using mapreduce. Lecture Notes in Computer Science, 7652:251-258.

[3] Yun, U.R., Yang, H., Yu, K.H. (2014) High utility itemset mining with techniques for reducing over-estimated utilities and pruning candidates. Expert Systems with Applications, 41(8):3861-3878.

[4] Sowan, B., Dahal, K., Hossain, M., et al. (2013). Fuzzy association rule mining approaches for enhancing prediction performance. Expert Systems with Applications, 40(17):6928-6937.