Analysis of Blast Furnace Data Based on Association Rules

Ze-kai CHENG¹, Qi ZHANG², Feng QIN³ and Li-qin ZHU⁴

School of Computer Science, Anhui University of Technology, Ma’anshan, Anhui 243032, China

Keywords: Blast furnace, Data mining, Association rules, Apriori algorithm.

Abstract. In this paper, taking the blast furnace temperature data as the data source, using association rules in data mining. Firstly, the practical significance of temperature data analysis of blast furnace in a steel plant is introduced. Secondly, the theory of data mining and association rules is given, detailed description of the principle of Apriori algorithm. This paper selects the temperature data of B blast furnace in a steel plant in September 2015 as the research object. Association Rule Mining for temperature data of blast furnace in a steel plant, and the association results are analyzed.

Introduction

Iron and steel is one of the most important materials in modern construction, blast furnace ironmaking plays a very important role in the economic development of our country. Blast furnace ironmaking process is very complex, it involves the conversion of various energy and matter, so it is very difficult to understand all kinds of phenomena in blast furnace[1].

In the production process, specific operations are based on the experience of experts and technical personnel to make judgments, therefore, there is a certain degree of unreliability, may affect the operation of the blast furnace. The application of data mining technology in the production process of blast furnace, the blast furnace in the production process of data analysis, the results are applied to the production process, assist experts and technicians to do the operation, it can effectively reduce the error rate, so as to improve the output of blast furnace and ensure the blast furnace.

Data mining, also called knowledge discovery in database. It is the process of extracting hidden, unknown, potentially valuable information and knowledge from large, incomplete, noisy, fuzzy, random, large data[2]. Simply say, data mining is to extract from a large number of data or mining useful knowledge, in the face of the current status of mass data and trace information, data mining is an important branch of the research and analysis, it is associated with the rules of the research aims mainly from a large quantity of data collection of the discovery, useful and property law.

By association rule mining, it can get useful information with potential value in implicit and massive data, the goal of association rules is to extract the most interesting patterns in an efficient way.

Theoretical Foundation

Association Rules

Association rules is an important research direction of data mining, the purpose of the association analysis is mining the relationships that are hidden in the data. Definition: Let I={i₁,i₂,...,iₘ} be an itemset. Let D, the task-relevant data, be a set of database transactions where each transaction T is not empty itemset such that T ⊆ I. Each transaction is associated with an identifier, called a TID[2]. Data set D={t₁,t₂,...,tₘ} is a collection of t₁,t₂,...,tₘ transactions.

Association rules can be described as: an implication of the form A ⇒ B, where A ⊆ I, B ⊆ I, A ≠ Ø,B ≠ Ø, and A ∩ B=Ø. where s is the percentage of transactions in D that contain A ∪ B. That is,

\[
\text{support } (A \Rightarrow B) = P (A \cup B)
\]
The rule has confidence $c$ in the transaction set $D$, where $c$ is the percentage of transactions in $D$ containing $A$ that also contain $B$. That is, 
$$
\text{confidence (} A \Rightarrow B \text{)} = P (B | A)
$$

The minimum support ($\text{minsup}$) and minimum confidence ($\text{minconf}$) are given by the user, if itemset $\text{sup} (X) \geq \text{minsup}$, then $I$ is a frequent itemset, the association rules that satisfy the minimum support and minimum confidence are called strong association rules.

**Apriori Algorithm**

Apriori algorithm is the first association rule algorithm, apriori employs an iterative approach known as a level-wise to search the database layer of itemsets, in order to form a rule, this process consists of the pruning and join. A set of items is referred to as an itemset. An itemset that contains $k$ items is a $k$-itemset. If the relative support of an itemset $I$ satisfies a prespecified minimum support threshold, then $I$ is a frequent itemset$[4][5]$.

Algorithm steps are as follows:

a) Setting minimum support $S$ and minimum confidence $C$.

b) Apriori algorithm uses candidate itemsets. Firstly, set of generating candidate options. If the candidate set is more than or equal to the minimum support, then candidate set is a frequent itemset.

c) In the first iteration of the algorithm, each item is a member of the set of candidate 1-itemsets, get all the support of item, the synthesis of 2-itemsets sets by using frequent 1-itemsets, because all nonempty subsets of a frequent itemset must also be frequent.

d) Re scan the database, get the candidate 2-itemsets, and then find out the frequent 2-itemset, and use these frequent 2-itemsets to generate candidate 3-itemsets.

e) Repeat scan the database, comparison with minimum support, generating higher level frequent itemsets, and the next candidate itemsets are generated from the set, until no new candidate itemsets are generated.

In this algorithm, two steps must be repeated: join and prune. The details are as follows.

a) Join. To find $F_k$, a set of candidate $k$-itemsets is generated by joining $F_{k-1}$ with itself. This set of candidates is denoted $I_k$. Let $l_1$ and $l_2$ be itemsets in $F_{k-1}$. The join, $F_{k-1} \otimes F_{k-1}$, is performed, where members of $L_{k-1}$ are joinable if their first $(k-2)$ items are in common.

b) Prune. $I_k$ is a superset of $F_k$, that is, its members may or may not be frequent, but all of the frequent $k$-itemsets are included in $I_k$. A database scan to determine the count of each candidate in $I_k$ would result in the determination of $F_k$.

**Data acquisition and Preprocessing**

**Data Acquisition**

In order to measure the temperature variation at the furnace wall, the B blast furnace in a steel plant, the detection of galvanic installed in the cooling wall and refractory materials at different heights, for temperature measurement in the blast furnace$[6]$. Blast furnace temperature detection points is shown in Figure 1.

![Figure 1. Schematic diagram of blast furnace.](image)
B blast furnace longitudinal with a total of 8 layer detection galvanic, tin the blast furnace 7,8,9,10,11,12,13,16 section of the cooling wall equipped with a number of temperature detection for detecting the temperature change of each section of the blast furnace, which 7,8,9,10,11,12 layer with six temperature detection galvanic, 13 and 16 layers are equipped with eight temperature detection galvanic. The name of each temperature detection is shown in Table 1.

<table>
<thead>
<tr>
<th>Layer</th>
<th>The name of thermocouple temperature detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>T07036 T07018 T07136 T07108 T07113 T07118 T07126 T07131</td>
</tr>
<tr>
<td>8</td>
<td>T08013 T08031 T08136 T08108 T08113 T08118 T08126 T08131</td>
</tr>
<tr>
<td>9</td>
<td>T09036 T09018 T09136 T09108 T09113 T09118 T09126 T09131</td>
</tr>
<tr>
<td>10</td>
<td>T10036 T10018 T10136 T10108 T10113 T10118 T10126 T10131</td>
</tr>
<tr>
<td>11</td>
<td>T11036 T11018 T11136 T11108 T11113 T11118 T11126 T11131</td>
</tr>
<tr>
<td>12</td>
<td>T12036 T12018 T12136 T12108 T12113 T12118 T12126 T12131</td>
</tr>
<tr>
<td>13</td>
<td>T13036 T13018 T13136 T13108 T13112 T13114 T13118 T13126 T13130 T13132</td>
</tr>
<tr>
<td>16</td>
<td>T16036 T16018 T16136 T16108 T16112 T16114 T16118 T16126 T16130 T16132</td>
</tr>
</tbody>
</table>

The data acquisition system collects data every minute, collecting temperature data for current time, and connect to the OPC server to store data on the SQL Server. The data used in this experiment is a steel B blast furnace data collected in 2015, September, the data acquisition is based on 68 temperature detection galvanic layer of 7, 8, 9, 10, 11, 12, each object collects 38931 temperature data.

**Data Pre-processing**

The data generated in the production process is vulnerable to noise, missing values and inconsistent data, low-quality data will result in poor quality analysis[7]. So through data preprocessing, the data analysis process will be more efficient and easier.

In the experiment, the method of quantile comparison chart is used to detect anomaly points, figure 2 is the quantile comparison chart of T08031 as the experimental object.

![Quantile comparison chart](image)

The black point in Figure 2 is the sample value in the original data, the red solid line is the normal distribution line, the red solid line contains a 98% confidence interval for the original data, the data not in the confidence interval will marked as an exception, and to be removed.

In order to meet the data requirement of Apriori algorithm, to transformation this data. Each data minus the previous data, if the result is greater than zero, the data is marked as 1, otherwise, the data is marked as 0, that is, if the data is 1, indicates that the state of the data is in a state of rising
temperature, if the data is 0, indicates that the state when the data is generated is a temperature drop state or a temperature unaltered state.

**Data Analysis and Conclusions**

Establishing "flow" analysis in SPSS Modeler[8]. "Flow" design is shown in Figure 3.

![Figure 3. Associated process graph.](image)

Using SPSS Modeler software for analysis, the resulting network analysis of the association analysis is shown in Figure 4.

![Figure 4. Associated network graph.](image)

Figure 4 shows the link between the points of the blast furnace, the thicker the line connecting two points in the network diagram represents the higher the correlation between two points, instead, show that the relevance is low.

The following conclusions can be drawn from Figure 4.

- The eighth, ninth, and twelfth layers are not strongly associated with other points, while the eleven, thirteen and sixteen layers are strongly associated with other points, indicating that these three layers of temperature change easily affect other points. So when observing the temperature index, special attention should be paid to the temperature change of these three layer.

- The points T07113, T11136, T111036, T16036 and T16108 can be found in the figure, which are more strongly associated with other points on the blast furnace, so when observing the temperature data, pay special attention to the above five points of temperature changes, it can be inferred that the other temperature points and the overall situation of the blast furnace.

- The line between point T07113 and point T16018, point T07113 and point T16036, point T16136 and point T16018, point T13036 and point T13136, point T13018 and point 13118, the line between point T11136 and T11036 is darker, there is a strong correlation between the two points. The strong link number of the point above is shown in Figure 5.
In the overall 38931 data, the number of strong links above the number of links have reached more than 7500, accounted for a large proportion of the whole, therefore, when we can observe the temperature data, the temperatures of the above-mentioned associated points are combined to be observed.

In the modeling process of SPSS Modeler software, set the minimum support to 4%, the minimum confidence is set to 70%, get the association rules, and get 84 rules of support, confidence, lifting degree, but considering that there are 47 association rules, the degree of lifting is less than 2, while the other 37 association rules are greater than 2, according to the concept of lifting degree, only 37 rules are valuable. Among them, 15 rules are selected as shown in Table 2.

<table>
<thead>
<tr>
<th>After item</th>
<th>Preceding item</th>
<th>Support %</th>
<th>Confidence %</th>
<th>Lifting</th>
</tr>
</thead>
<tbody>
<tr>
<td>T13136</td>
<td>T12113 T13036 T13132</td>
<td>5.027</td>
<td>77.568</td>
<td>2.672</td>
</tr>
<tr>
<td>T13036</td>
<td>T13136 T16108 T12031</td>
<td>5.435</td>
<td>79.679</td>
<td>2.434</td>
</tr>
<tr>
<td>T13036</td>
<td>T13136 T13132 T12108</td>
<td>5.743</td>
<td>78.354</td>
<td>2.394</td>
</tr>
<tr>
<td>T11136</td>
<td>T10108 T12108 T11036</td>
<td>10.043</td>
<td>80.409</td>
<td>2.34</td>
</tr>
<tr>
<td>T11036</td>
<td>T10108 T12031 T11136</td>
<td>9.614</td>
<td>80.924</td>
<td>2.19</td>
</tr>
<tr>
<td>T11036</td>
<td>T13036 T10108 T11136</td>
<td>6.547</td>
<td>80.659</td>
<td>2.182</td>
</tr>
<tr>
<td>T11036</td>
<td>T13136 T10108 T11136</td>
<td>6.288</td>
<td>80.474</td>
<td>2.177</td>
</tr>
<tr>
<td>T10126</td>
<td>T10113 T10118 T10018</td>
<td>4.367</td>
<td>70.059</td>
<td>2.53</td>
</tr>
<tr>
<td>T09131</td>
<td>T10113 T10118 T10136</td>
<td>4.513</td>
<td>70.746</td>
<td>3.093</td>
</tr>
<tr>
<td>T09118</td>
<td>T09108 T09113 T09018</td>
<td>4.485</td>
<td>70.046</td>
<td>2.597</td>
</tr>
<tr>
<td>T08108</td>
<td>T07126 T07108 T08131</td>
<td>5.255</td>
<td>86.119</td>
<td>5.443</td>
</tr>
<tr>
<td>T07131</td>
<td>T08108 T07126 T07108</td>
<td>5.41</td>
<td>86.705</td>
<td>5.052</td>
</tr>
<tr>
<td>T07131</td>
<td>T07126 T07108 T07136</td>
<td>5.407</td>
<td>86.318</td>
<td>5.029</td>
</tr>
<tr>
<td>T07126</td>
<td>T08108 T07131 T07118</td>
<td>5.315</td>
<td>86.129</td>
<td>5.134</td>
</tr>
<tr>
<td>T07118</td>
<td>T08108 T07126 T07108</td>
<td>5.41</td>
<td>86.04</td>
<td>4.913</td>
</tr>
<tr>
<td>T07108</td>
<td>T08108 T07126 T07131</td>
<td>5.356</td>
<td>87.578</td>
<td>5.177</td>
</tr>
</tbody>
</table>

Table 2 shows the results of apriori algorithm for all points, the data in Table 2 indicates that the after term is associated with the preceding item, after the after item change, the preceding item may also change, the support degree indicates the probability that the preceding item and after item occur simultaneously in the overall transaction, confidence indicates the probability of occurrence.
of after term in the case where the preceding term occurs, lifting degree represents the value of the information mined[9]. Therefore, From the table can be seen that the confidence of the results of the vast majority are greater than 75%, and the degree of lifting is greater than 2, so that we can know, the results obtained by the Apriori algorithm are valuable.

The 37 rules are summarized in Table 3.

<table>
<thead>
<tr>
<th>After item</th>
<th>Preceding item</th>
</tr>
</thead>
<tbody>
<tr>
<td>T13136</td>
<td>T12113, T13036, T13132, T16108</td>
</tr>
<tr>
<td>T13036</td>
<td>T13136, T16108, T12031, T13132, T12108</td>
</tr>
<tr>
<td>T11136</td>
<td>T10108, T12108, T11036, T12031, T10036, T11036</td>
</tr>
<tr>
<td>T11036</td>
<td>T12031, T11136, T12108, T10108, T13036, T10018</td>
</tr>
<tr>
<td>T10126</td>
<td>T10113, T10118, T10018, T09118</td>
</tr>
<tr>
<td>T09131</td>
<td>T10113, T10118, T10136</td>
</tr>
<tr>
<td>T09118</td>
<td>T09108, T09113, T09018</td>
</tr>
<tr>
<td>T08108</td>
<td>T07126, T07108, T08131</td>
</tr>
<tr>
<td>T07131</td>
<td>T08108, T07126, T07108, T07136, T07118</td>
</tr>
<tr>
<td>T07126</td>
<td>T08108, T07131, T07118</td>
</tr>
<tr>
<td>T07118</td>
<td>T08108, T07126, T07108</td>
</tr>
<tr>
<td>T07108</td>
<td>T08108, T07126, T07131, T07136, T07131, T08131</td>
</tr>
</tbody>
</table>

The following conclusions can be drawn from table 3:
- When the blast furnace temperature detection point of some anomalies, staff not only need to take into account the situation, but also be based on the data in the table to see the point with the relevant temperature point of the actual situation, it is also possible that the exception that occurs at that point is caused by an anomaly at the associated point, staff should take these circumstances into account, so as to make the corresponding operation.

Conclusion

In this paper, we select the temperature data of a steel plant B blast furnace in September 2015 as the research object, the association rules are used to find out the association between data, which uses the association graph and Apriori algorithm to carry on the experiment, finally, the association between data was analyzed. Through the conclusions of this experiment, so that experts and technical staff can better understand the status of blast furnace temperature, can effectively reduce the error rate, thereby increasing the production of blast furnace and ensure the blast furnace line. This work was sponsored by a grant from major projects supported by natural science in Anhui province in 2014 (KJ2014ZD05).

References


