Research on Overall Equipment Effectiveness for Cigarette Equipments Based on Data Mart and Data Mining

Li Lin, Huang Xiu, Wan Yunfei and Wang Guangchao

ABSTRACT

In the basis of building data mart, this paper uses data mining clusters and association rule algorithms from people, machines, materials, regulations, environments and measurement systems to excavate the key influencing factors which influence the overall equipment effectiveness for cigarette equipments. It has a great significance for equipment management decision support, optimizing the equipment management assessment mechanism and enhancing the level of equipment management.

INTRODUCTION

Overall Equipment Effectiveness (OEE) is one of the most important contents for TnPM, which has become the basic tool for modern lean production theory.

At present, OEE analysis is limited to the numerical calculation for cigarette equipment’s, which is non-scientifically, non-effectively and non-reasonably for analyzing the key factors affecting on OEE for cigarette equipment’s. Business information systems which generated data cannot be used for good and live to mining the hidden value of data to achieve the preservation and appreciation for data.

Based on constructing data mart [1], this paper analyzes the key factors affecting on OEE from the six dimensions of people, machines, materials, regulations, environments and measurement systems. It has a great significance for equipment management decision support, optimizing the equipment management assessment mechanism and enhancing the level of equipment management.

Li Lin¹, Huang Xiu²
China Tobacco Shandong Industrial Co., Ltd.¹-²
Zao zhuang, China
baolinlee@163.com

Wan Yunfei³, Wang Guangchao⁴
China Tobacco Shandong Industrial Co., Ltd.³-⁴
Zao zhuang, China
baolinlee@163.com
CONCEPT

Data Encryption Standard

At present, several of information systems in the use in Teng Zhou cigarette factory has a total for more than 20. Data is scattered stored and information resources cannot be shared so as to forming a data island. Moreover, in the process of system implementation, all kinds of systems do not have an agreement on data type and storage format, lacking of data storage uniform standards. Which leading to multi-systems to acquisition the same index data of the same equipment, but data collection is non-comparable and the credibility of the data decreased.

To form an analysis-oriented environment, we must ensure the accuracy and completeness of data firstly. To develop a unified data coding standard is the basis of building a data mart for data mining [2].

OEE data mart is based on the coding principle of data classification, identification and coding uniqueness, rationality, expansibility, conciseness, adaptability and normative to be used for tobacco, tobacco materials, spare parts, finished cigarettes, semi-finished products, cost center, work center, production version, team, machine, supplier, customer, account, fixed asset, production order type, fund center, commitment items, propaganda and promotional items to construct classification standard, description standard and coding principle.

Data Storage and Retrieval

The main source of Teng Zhou cigarette factory data include three parts. Firstly, data for the ring network collecting, secondly, data from MES, ERP and other information systems, thirdly data for manual input. Insert several of data into OEE data mart by ETL [3] tools. ETL means data extraction [4]; transmission and loading process. Extract from the business system database platform terminal all kinds of business data according to the specified frequency, unified storage in the data mart server. The data storage is designed by the combination of relational database and unstructured database. The relational database mainly stores the business data integrated by each business system and the related rule data. The unstructured data is used to store data such as various of rules, policies and regulations.

Data Mart

Data mart is a subject-oriented, departmental data warehouse. Data analysis in the data mart [5] must be effective in the realization of data extraction based on the establishment of data model [6]. Data tables are stored as fact tables and dimension tables in data marts. It appears as star structures on logical models. The fact table [7] is at the center of the star structure. Which is a set of granularity attribute and measure value. The dimension table [8] is a set of the data. Both the fact table and the dimension table are to construct multi-dimensional organization structure to form the data cube.

OEE theme logic model of Teng zhou cigarette factory for cigarette equipment’s as shown in figure 1:
DATA MINING FOUNDATION FOR OEE

General Idea for OEE Data Mining

OEE analysis model for cigarette equipment’s was carried out based on Tengzhou cigarette factory basic data and use fishbone diagram as the influencing factor analysis method [9], data mining as the tool for analysis.

By analyzing the relationships among people, equipment’s, materials, regulations, environments, measurement systems, we find out which factors are strongly correlated and which are weakly correlated. One is to use association rules mining algorithm to find out the reasons for low OEE based on the simultaneous appearance of several influential factors. For example, which factors simultaneously will lead to high failure rate for equipment, and which factors will lead to a high rate of product unqualified? The second is to discover the strategy of improving OEE by sequential pattern mining [10], such as how to arrange the process (equipment maintenance cycle and spare part replacement cycle), and improve the time utilization ratio of the equipment. Equipment maintenance cycle and spare parts replacement cycle to provide a reference. Which equipment to produce which brand will make OEE is the best to provide a guide for the production plan.

OEE Influencing Factors

OEE data mining mainly includes six dimensions: people, equipment’s, materials, regulations, environments and measurement systems. The fishbone diagram as shown in figure 2:
DATA MINING ANALYSIS FOR OEE

In this paper, data mining is carried out on the basis of constructing the theme data market of OEE for cigarette equipment. The "knowledge model" of people, equipment’s, materials, regulations, production environments, and measurement systems are acquired as the knowledge base and method library of "factor analysis" and "judgment analysis", together with the time series analysis, this system will greatly improve the forecasting level and make the forecast more in line with the actual situation. It can meet the actual demand to arrange the equipment maintenance and replacement, effectively control the production cost, and comprehensively improve OEE for cigarette equipment’s. Enterprise OEE control analysis structure as shown in figure 3:

Data Preprocessing

Data preprocessing is mainly to organize data from the existing enterprise information system based on the data mining algorithm model which abandon some of the attributes associated with the mining target to provide clean, accurate and more targeted data for the data mining kernel algorithm, thus reducing the number of data mining core processing capacity to improve the efficiency of mining, to improve the knowledge discovery starting point and the knowledge accuracy.

Data preprocessing mainly include two tasks: ① encoding the concept tree; ② with its code instead of the contents to constitute the encoding database.

For each concept level tree the continuous integer encoding, and then coding each tree node layer by layer, the code of the child node is the parent node code and
child nodes in the sub tree in the combination of the serial number. The hierarchy concept of the device synthesis efficiency theme as shown in figure 4:

![Figure 4. OEE Concept hierarchy tree for cigarette equipment's.](image)

OEE hierarchy coding tree which corresponding to OEE Concept hierarchy tree for cigarette equipment as shown in figure 5:

![Figure 5. OEE hierarchy coding tree for cigarette equipment's.](image)

Then use the code in the database instead of coding to form the database, after the mining association rules are encoded in this database. And then the database entry with the code instead, constitutes the encoding database T [1], as shown in table 1 below. Subsequent mining of multi-layer association rules is carried out in this coded database.

<table>
<thead>
<tr>
<th>Table 1. Concept coding data base T [1].</th>
</tr>
</thead>
<tbody>
<tr>
<td>TID</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>T1</td>
</tr>
<tr>
<td>T2</td>
</tr>
<tr>
<td>T3</td>
</tr>
<tr>
<td>T4</td>
</tr>
<tr>
<td>T5</td>
</tr>
<tr>
<td>T6</td>
</tr>
<tr>
<td>......</td>
</tr>
</tbody>
</table>

**OEE Data Mining**

The data mining efficiency of the equipment is mainly based on the data mining clustering algorithm and the association rule algorithm.

(1) EM algorithm is used to classify the data into two main categories.

The expectation maximization algorithm is computed in two steps. The first step is to compute the expectation (E), which is the expected value that the hidden variable can be observed to include the same as the maximum likelihood. The other step is to maximize (M), that is, maximizing the maximum likelihood expectation found at step E to calculate the maximum likelihood estimate of the parameter. The parameters found on step M are then used for another E-step calculation, and the process continues alternately. After all the influencing factors are taken into account, the expectation value and the maximum likelihood estimation value are high, and the others are not.
From the EM algorithm, we analyze the optimal combination of man-machine-material-loop method to maximize the efficiency of the integrated operation of equipment. To analysis the reasons of OEE lower from others data so as to improve the efficiency of integrated equipment to provide advice and reference for leader.

Although the EM algorithm can guarantee convergence to a certain maximum, it may be only a local maximum, not a global maximum. In order to obtain the global maximum, the whole process is repeated several times with different initial guess values, and the largest parameter configuration is selected.

(2) Multi-level FP-tree association rules mining algorithm respectively, the two types of data were associated analysis.

In the OEE analysis for cigarette equipment, the minimum support is 5 for layer 1, the minimum support is 3 for layer 2, and the minimum support is 2 for layer 3, the coding transaction table generated by part data as shown in table 2:

<table>
<thead>
<tr>
<th>TID</th>
<th>Item encoding</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>{111,121,211,221}</td>
<td>1</td>
</tr>
<tr>
<td>T3</td>
<td>{111,211,222,323}</td>
<td>1</td>
</tr>
<tr>
<td>T4</td>
<td>{112,122,221,411}</td>
<td>1</td>
</tr>
<tr>
<td>T6</td>
<td>{111,121}</td>
<td>1</td>
</tr>
<tr>
<td>T8</td>
<td>{111,122,211,221}</td>
<td>1</td>
</tr>
<tr>
<td>T10</td>
<td>{211,321,524}</td>
<td>1</td>
</tr>
<tr>
<td>T11</td>
<td>{313,411,513}</td>
<td>1</td>
</tr>
</tbody>
</table>

The class frequent pattern tree generated by the encoded transaction table of table 2 as shown in Figure 6:

After all the influencing factors are taken into account, the expectation value and the maximum likelihood estimation value are high, and the others are not. The results of cumulative analysis of OEE for cigarette equipment as shown in figure 7:
By calculating the probability of several attributes occurrence at the same time, and in the condition that when one attribute occurrence several other properties will also occurrence to determine the key factors affecting on OEE for the cigarette equipment.

The association rule analysis of OEE for cigarette equipment as shown in figure 8:

**OEE Data Mining Results**

By mining data from January 2015 to September 2016, we have the following rules:

Replacement of the brand will lead to OEE decline for the cigarette equipment. (The probability of occurrence is 1)

3 #, 11 # cigarette equipment’s are the most suitable for the production Hade door (soft), when the OEE is the highest for cigarette equipment. (The probability of occurrence is 0.99)

When humidity is not up to standard in the workshop, it is easy to cause wrapping paper quality defects. (The probability of occurrence is 0.94)

Based on the results of the data mining analysis, we make the following decisions for the leadership:
<table>
<thead>
<tr>
<th>NO.</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>It is recommended that to minimize the frequency of brand replacement for the same equipment.</td>
</tr>
<tr>
<td>2</td>
<td>It is recommended that 3 #, 11 # cigarette equipment’s production Had door (soft).</td>
</tr>
<tr>
<td>3</td>
<td>It is recommended to establish the workshop temperature and humidity assessment management regulations to reduce the temperature and humidity is not up to standard, which resulting in product quality defects.</td>
</tr>
</tbody>
</table>

**CONCLUSION**

Based on the basis of standardization of the data for each business information system, this paper constructs OEE theme data mart and uses data mining method to compare, analyze and explore the potential value of the data. It can improve the ability of production process control, decision analysis ability, improve enterprise OEE has very important practical significance, with strong maneuverability.

**REFERENCES**


[4] Hong Cheng, Xifeng Yan, Jiawei Han and Chih-Wei Hsu, Discriminative Frequent Pattern Analysis for Effective Classification [C], Proceedings of the 2007 IEEE International Conference on Data Engineering (ICDE’07), Istanbul, Turkey, April 2007.


[6] Hong Cheng, Xifeng Yan, Jiawei Han, and Philip S. Yu, Direct Discriminative Pattern Mining for Effective Classification [C], Proc. 2008 Int. Conf. on Data Engineering (ICDE’08), Cancun, Mexico, April 2008.


