Quality Analysis Method for Hot Strip Rolling Based on Data Mining Technology

Zhen Hou[1, a], Hai Gao[2, b], Hua Chen[3, c]

[1]School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

[2]School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

[3]No.12 Rujiang West Road, Mawei Hi-tech Development Zone, Fuzhou Fujian 350015, China

[a]g20158510@xs.ustb.edu.cn, [b]gaohai@ustb.edu.cn, [c]ch052800@163.com

Keywords: hot rolling of strip steel, data mining technology, correlation analysis, quality analysis, control parameters.

Abstract. The hot rolling of strip steel is a continuous, multi-staged, complex production process, and there are about one hundred of control parameters which are directly related to the quality of strip steel products in this complicated process. Researches have shown that, according to the industrial characteristics of hot strip rolling, using data mining technology to extract the useful, potential and ultimately understandable process knowledge, and get the corresponding relationship between the strip steel quality defects and control parameters. It can quickly locate the causes of strip steel quality problems, and find out the key control parameters to make adjustments, improve production efficiency and production quality, and reduce economic losses. It provides a scientific and accurate way to analyze the quality problems of hot rolled strip steel products.

1. Introduction

Strip steel is a narrow but long steel plate produced by various steel mills in order to meet the needs of different industrial sectors for the production of various types of metal or mechanical products, and the width is within 1300mm,and the length is slightly different according to the size of each volume. generally speaking, strip steel can be divided into ordinary strip steel and high quality strip steel according to the material type. In terms of processing method, it can also be divided into hot-rolled strip steel and cold-rolled strip steel[1].

Data mining technology is a process to extract the implied, unknown, potentially valuable information and knowledge from the vast, messy, vague, random and practical application data[2].It is a new and deep-seated information processing technology, and its main purpose is to extract, transform, analyze and model a large number of data in the data set, and to obtain the key data information which helps to make decisions. The general data mining process is shown in Fig. 1
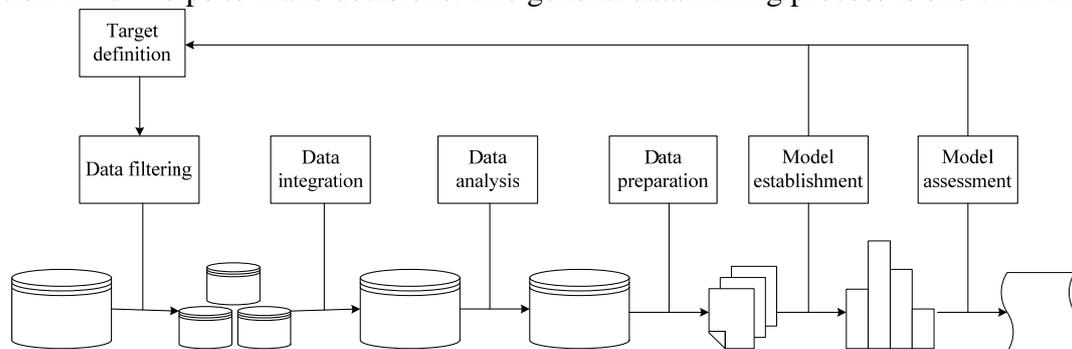

Fig. 1 Data mining process

This article is based on a data set of the strip steel with 4949um thickness specification from a steel mill by using the IbaAnalyzer software as a platform, and the improved factor analysis method, the Apriori algorithm and the FP-growth algorithm as methods to mine the knowledge, rules and

patterns in the data set, in order to provide the evidence for decision-making when trying to find the main control parameters that affect the thickness of strip steel.

2.   The introduction of analysis method for strip steel thickness quality index

2.1 The improved factor analysis method.

Factor analysis[3] is a multivariate statistical analysis method which is aimed at finding the potential factors who hiding in the multivariate data and can not be directly observed but influence or dominate measurable variables, and furthering to estimate the impact of potential factors on measurable variables. The basic idea is to start from the analysis of the relationship between multivariate data to find a few independent potential factors that dominate this correlation and to help find some kind of objective regularity hidden between in the original variable. However, these potential factors are not measurable in the production process of hot strip rolling, and they can not be used as control parameters for technical personnel to observe and adjust intuitively. This article puts forward an improved factor analysis method based on the traditional factor analysis method, we can visually get the main control observation variables which influence the strip steel thickness index, and reduce the dimension of the state space matrix.

The dimension reduction steps of state space data set based on improved factor analysis method is as follows:

(1)To extract the principal component factors and to reduce dimensions.

The first, second, ..., p^{th}(p≤n,n is the number of control parameters)potential factor which impacts target parameter can be obtained by analyzing the original data through SAS statistical analysis software with Principal Component Analysis[4,5] method. The load l_{ij} of each control c_j variable on the pth principal component factor m_i can be obtained by the following formula:

$$l_{ij} = p(c_j, m_i) = \sqrt{\lambda_i e_{ij}} \quad (i = 1,2,...,p, j = 1,2,...,n) \tag{1}$$

(2)To rotate the potential factors.

However, when the load distribution of some potential factors in each control parameter is uniform. It is not easy to see directly which control parameter has the strongest relationship with the potential factors, also it is not easy to give the potential factors a reasonable explanation meaning. In this case, the factorial rotation method is used in which the initial potential factors are transformed into a new set of potential factors by using a linear transformation. The new potential factors give an absolute value of the factor load on each control variable to 0 or 1 polarization, so it is clear that each potential factor has the greatest effect on a control parameter, and then gives a more reasonable explanation of each potential factor.

(3)To evaluate the influence of the key control parameters.

The influence of the key control parameters(w_j), that is the effect weight of the key control parameters on the strip steel thickness quality index, is numerically equal to the product of the most significant load absolute value of the kth potential factor to the jth control parameter and the contribution rate of the kth potential factor[6].The formula is as follows:

$$w_j = |l_{kj}\lambda_k| \quad (k = 1,2,...p, j = 1,2,...,n) \tag{2}$$

2.2 FP-growth algorithm.

The FP-growth algorithm is an association analysis algorithm proposed by Jiawei Han et al. in 2000.It compresses the data set that providing frequent item sets into a frequent pattern tree (FP-tree),which still retains the item-related information. The FP-tree is a special prefix tree which consisting of frequent item headers and item prefix trees[7].

As we all know, the Apriori algorithm needs to scan the data set many times and generate a large number of candidate frequent item sets before generating the final frequent item sets, which not only spends more time, but makes space more complicated, so the performance of mining frequent item sets tends to be low.However, the Apriori algorithm has a very important nature: all non-empty subsets of frequent item sets must also be frequent. That is to say if an element is infrequent, then its supersets are also infrequent[8,9].Based on this important property, Jiawei Han proposed the FP-Growth algorithm. The FP-Growth algorithm can obtain the frequent item sets by using the FP-tree

structure to only scan the whole data set twice, thus speeding up the whole mining process. Fig. 2shows the general form of FP-tree.
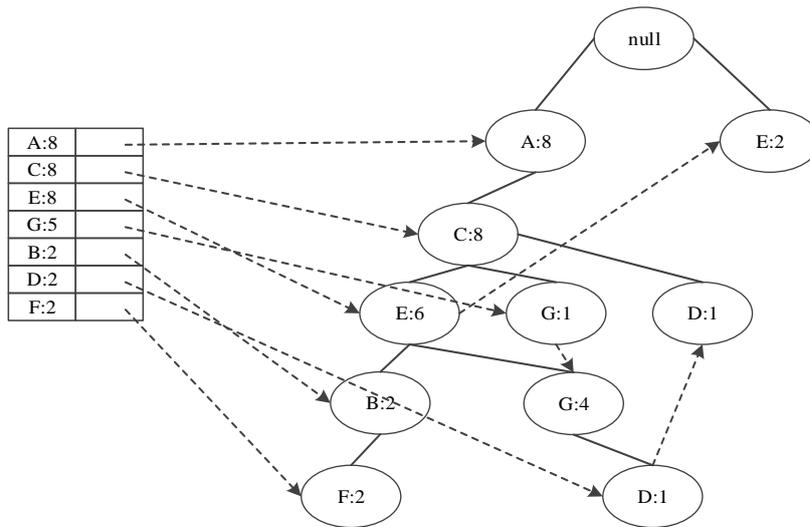

Fig. 2 FP-tree structure

2.3 Correlation analysis model of strip steel thickness deviation based on FP - growth algorithm.

It is an iterative process to build model using machine learning algorithm. adjusting the model parameters and optimizing the model, the problem can be solved better.The working process of the FP-growth algorithm is as follows:

(1)To build the FP tree. To build an FP tree, you need to scan the original data twice. Count all element entries for the first time.

(2)To determine the minimum support.

(3)To mine frequent item sets using FP trees and minimum support parameters.

(4)To determine the minimum confidence.

(5)To discover association rules. Since the FP-growth algorithm can only find frequent item sets more efficiently and can not be used to discover association rules, the association rules are discovered by using Apriori algorithm for the frequent item sets that have been found [10].

3. Experimental results

It is necessary to understand the production process of the hot strip rolling before analyzing and mining data, only in this way can we understand the influence factors of strip steel thickness quality index properly. The production process of hot rolling strip includes phosphorus removal, fixed width, rough rolling, flying shear, finish rolling, cooling, coiling and other processes [11].The flow chart is shown in Fig. 3.
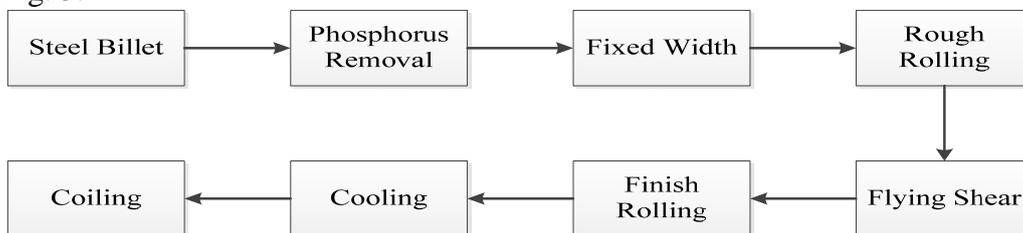

Fig. 3 Production process of hot strip rolling

In the production process of hot strip rolling, strip steel have many quality indexes, and the thickness accuracy index is an important aspect of the size and shape accuracy in strip steel product quality index. It is not only directly related to the quality and economic benefit of strip steel products, but also determines the use and application field of strip steel products. Therefore, it is of great significance to ensure the thickness accuracy of strip steel products in production process.

3.1 The analysis of influence factors on strip steel thickness.

There are many factors that cause the thickness deviation of strip steel, such as the change of the rolling temperature, the change of the inlet thickness, the chemical composition of the steel steel, the

difference of heating temperature, the watermark, the zero drift of the vertical roller, the shape of vertical roller, the rolling force of finishing mill, the rolling force of roughing mill, the roll gap of finishing mill, the roll gap of roughing mill, the looper angle, the looper tension and the control model of thickness. According to the field experience of hot strip rolling production and combine with the analysis and summary of the data set,the key issue of thickness control is that:

(1)The thickness of the whole strip steel is ultra-thick or ultra-thin.

(2)The head and tail of strip steel are ultra-thick.

(3)The thickness of strip steel fluctuates greatly and is uneven when changing specifications.

(4)The rolling accuracy is low at the initial stage when changing rollers, and the parameters differ greatly from the previous roller cycle.

(5)The instability of rolling force control leads to a large error of thickness in some strips steel [12,13].

3.2 To obtain and preprocess the data set.

Data preprocessing is the preparatory work of data mining technology, the quality of the modeling data directly affects the accuracy of the model. Generally speaking, the data set has two salient features: one is that the data is dirty and the other is that the sample size is large. Therefore, in order to improve the accuracy, effectiveness and scalability of modeling data, the data set needs to be processed as the following steps:

(1)To obtain the data set.

In the production process of hot strip rolling, the steel billet needs to be continuously rolled through seven finishing mills. Since the length of the strip steel constantly changes and the linear velocity of rollers also changes, the number of observed values measured by each rolling mill is different. For example, the first finishing mill measured 1250 groups of data, including F1 FORCE, F1 GAP, F1 SPEED, F1 CURRENT, LP1 ANGLE, LP1 TENSION, and the second finishing mill measured 1350 groups of data, including F2 FORCE, F2 GAP, F2 SPEED, F2 CURRENT, LP2 ANGLE, LP2 TENSION, and the third finishing mill measured 1450 groups of data, including F3 FORCE, F3 GAP, F3 SPEED, F3 CURRENT, LP3 ANGLE, LP3 TENSION......In order to ensure that the measured values in each group of data correspond to the same position of the strip steel, In this article, 50 strips of steel are selected, and 400 sets of data are extracted from each strip in percentiles, thus ensuring that each observation data in each set is measured from the same location.

(2)To preprocess data and obtain modeling data.

The data set contains 33 features, the data set is huge and some data are invalid. The following are the steps of data preprocessing process:

①To clean data.

The purpose of data cleaning is to eliminate or reduce noise data or to handle missing values and abnormal value. The smoothing technique can be used to eliminate or reduce noise data. The missing values can be interpolated by Lagrange interpolation method, or they can replaced by the most likely value according to statistics, or directly deleted. This article uses the Lagrangian interpolation method to interpolate the missing values and delete the abnormal value. So data cleaning will be helpful to reduce the confusion of machine learning.

②Correlation analysis.

Many features are unrelated to problems. For instance, strip steel inlet coil number, strip steel export volume number, test time, material number and other characteristics do not affect the strip steel thickness. In addition, some attributes are redundant possibly. If these attributes are included in the data set, the machine learning steps will be slowed down or may be misguided. Therefore, it is necessary to analyze the data in advance to remove irrelevant or redundant features in the machine learning process. Through the study of the parameters of the strip steel, it was decided to select the control parameters of the finish rolling stage as the original data which affect the strip steel thickness. Such as F1-F7 FORED, F1-F7 GAP, F1-F7 SPEED, F1-F7 CURRENT, LP1-LP6 ANGLE, LP1-LP6 TENSION.

③Dimension reduction based on improved factor method.

The first, second, ..., p^{th} (p≤n,n is the number of control parameters) potential factor which impacts the strip steel thickness quality index can be obtained by analyzing the original data through using SAS statistical analysis software with Principal Component Analysis method. The statistical analysis results as shown in Table 1.

Table 1. Factor analysis results

Factor model						
Variables	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
speed5	0.90297	-0.25703	0.15040	0.29816	-0.00453	0.02466
speed6	0.90077	-0.29321	0.10392	0.28812	-0.00310	0.02177
speed4	0.89992	-0.19421	0.23812	0.29216	-0.00497	0.02969
speed7	0.88560	-0.33191	0.07963	0.29345	-0.00043	0.02152
speed3	0.87986	-0.12469	0.34170	0.27902	-0.00151	0.03790
...	...	…	...	...	...	...
tension4	-0.06870	0.32036	-0.34433	0.25633	-0.04484	0.25611
tension5	-0.08824	0.31580	-0.29062	0.27146	0.60790	-0.36970
angle1	-0.11761	-0.33092	0.40938	-0.10370	0.54100	0.20466
tension6	-0.03485	0.24190	-0.47564	0.14703	-0.10675	0.48197
tension1	-0.17318	0.28296	0.04632	0.24391	-0.49599	-0.56930

Table 1 shows the load matrix between the potential factors and the control parameters. However, we can see that the load distribution of some potential factors in each control parameter is uniform. It is difficult to see which control parameter is influenced greatly by a potential factor directly, and thus it is not easy to give the potential factor a reasonable explain meaning.

In this article, the variance maximum orthogonal factor rotation method is used to change the position of the coordinate axis and reallocate the variance ratio explained by each factor, which makes the structure of the factor simpler and easier to explain. The relationship between the potential factors and the control parameters after factor rotating is shown in Table 2:

Table 2. Factor rotation analysis results

Rotation factor mode						
Variables	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
gap2	0.98765	0.04992	0.08450	-0.06482	0.03137	0.03400
gap5	0.98152	0.12729	0.06331	-0.04624	0.00806	0.03498
gap4	0.97351	0.08392	0.05037	-0.05716	0.01450	0.03163
gap3	0.96778	0.09613	-0.00599	-0.08183	0.04759	0.02808
gap6	0.95339	0.22890	-0.06211	-0.07627	0.02516	0.04943
...	...	...	...	...	...	...
tension6	-0.00621	-0.06635	-0.00668	-0.70430	-0.14949	-0.16851
tension5	0.07042	-0.07930	0.04822	-0.18842	0.84979	-0.02584
angle5	-0.02322	0.15874	0.16045	0.58488	-0.61688	-0.07065
tension1	0.08913	-0.06135	0.23891	-0.01901	0.07204	0.81698
angle1	-0.13855	0.04337	0.17243	0.46450	0.10460	-0.59895

In the process of factor analysis, the principle of 0.5 is generally used, that when the factor load value is greater than or equal to 0.5,then it is considered that the potential factor dominates the control variable. In this article, in order to select the most significant potential factor in the ability to dominate, the maximum factor load in k^{th} column is selected, that means the kth potential factor has the most significant load value on the jth control parameter in table 2,and this control parameter can be regarded as the kth key control parameter. In this way, all key control parameters are available. In addition, the effect weight of key control parameters are obtained by formula (2).The result in shown in Table 3.

Table 3. The effect weight of key control parameters

Control parameters	Load	Contribution rate	Weights
Gap2	0.98765	0.33350	0.98770
Speed2	0.93679	0.22680	0.93680
Force1	0.94687	0.17520	0.94690
Angle3	0.81480	0.05750	0.81480
Tension5	0.84979	0.03440	0.84980
Tension1	0.81698	0.02820	0.81700

④To transform data and to construct modeling data.

In the finish rolling step of the production process of hot strip rolling, the type of the data set measured by sensors is continuous. In order to meet the requirements of the FP-growth algorithm for modeling data, the continuous data need to be discretized. In this article, each control parameter is divided into A, B and C three categories by K-means clustering algorithm as shown in Table 4.According to the strip steel thickness index in the production process as shown in Table 5,the strip steel thickness error values are divided into 1,2,and 3 three levels in accordance with (0,50), (50,70), (70,500).

Table 4. Classification rules of key control parameters

Control parameters	A	B	C
Angle3	0	1776.893036	2248.031506
Force1	0	1566.600924	1842.999949
Gap2	0	1615.971681	1705.612809
Speed2	0	1778.264219	2142.044949
Tension1	0	6257.654478	7551.560564
Tension5	0	9001.165464	13946.839600

Table 5. Quality index of steel strip in actual working process of steel mills

Thickness(mm)	Thickness deviation(mm)	Percentage(%)
<2.0	±0.030	95.40
2.0-4.0	±0.040	95.40
4.0-6.0	±0.050	95.40
>6.0	±0.080	95.40

3.3 To mine frequent item sets based on FP-growth machine learning algorithm.

To get the frequent item sets,the association rules and the corresponding confidence that affect the strip steel thickness based on the Python development platform. The strip steel thickness deviation is taken as the output variable and F1-F7 FORCE, F1-F7 GAP, F1-F7 SPEED, F1-F7 CURRENT, LP1-LP6 ANGLE, LP1-LP6 TENSION are considered as input variables.

The first step is to use the FP-growth algorithm to mine the frequent item sets of the modeling data.To set up the minimum support is 0.5[14],and the frequent item sets that satisfy the minimum support are shown in Table 6.

Table 6. Frequent Item Sets

Frequent item sets	Counts	Support
Angle3_C, Speed2_A	7363	0.7363
1, Angle3_C	7398	0.7398
1, Speed2_A	6073	0.6073
1, Angle3_C, Speed2_A	5727	0.5727

The second step,the Apriori algorithm is used to analyze the association rules of the frequent item sets, the minimum confidence is set up to 0.75[15,16]. The results are shown in Table 7.

Table 7. Association Rules

Association rules	Confidence
Force1_A,Gap1_A-->1	0.968515431
Speed2_A-->Angle3_C	0.884118674
Angle3_C,Tension1_B-->1	0.818854569
Speed2_A-->1	0.784159159
Angle3_C, Speed2_A-->1	0.863848134

Association rule 1:

If Force1 belongs to [0,1566.6] and Gap1 belongs to 【0,1615.97】,

Then the level of strip steel thickness quality index is 1,

The thickness deviation belongs to [0,50],

The confidence is 96.85%.

Association rule 2:

If Angle3 >= 2248 and Tension1 belongs to 【6257.65, 7551.56】,

Then the level of strip steel thickness deviation is 1,

The thickness deviation belongs to [0,50],

The confidence is 81.89%.

Association rule 3:

If Speed2 belongs to 【0, 1778】,

Then the level of strip steel thickness deviation is 1,

The thickness deviation belongs to [0,50],

The confidence is 78.42%.

Association rule 4:

If Angle3 >= 2248 and Speed2 belongs to 【0, 1778】,

Then the level of strip steel thickness deviation is 1,

The thickness deviation belongs to [0,50],

The confidence is 96.85%.

Through the above results, it can be seen that the rolling force, roller gap, linear velocity of roller, looper angle and looper tension have important influence on the strip steel thickness, it mainly focus on a certain range of the characteristic value. Therefore, in order to achieve the strip steel thickness error in the range of[0,50],the staff can set the values of F1-F7 FORCE, F1-F7 GAP, F1-F7 SPEED, and F1-F7 SPEED and other key control parameters scientifically and reasonably according to the frequent item sets and association rules in the actual production process of hot strip rolling.

4. Conclusion

By preprocessing the original data derived from the production process of hot strip rolling, including data cleaning, correlation analysis, data dimension reduction, data transforming, the modeling data conforming to the modeling requirements are obtained. Then the FP-growth algorithm is used to build FP-tree to find the frequent item sets, finally to find association rules from frequent item sets by using Apriori algorithm. So it provides good decisions for technical personnel to apply these rules in the actual production process to improve the pass rate of strip steel products.

References.

[1]. Information on: http://baike.baidu.com/item/带钢

[2]. Information on: http://baike.baidu.com/item/数据挖掘技术/910318

[3]. Information on: http://baike.baidu.com/item/因子分析

[4]. Wu S,Wei Z B.Structure damage identification based on residual error of principal component. Journal of Highway and Transporta-tion Research and Development.Vol.28 (2011) No.7,p.113.

[5]. Wei-wei LIU,Hu LP .J Chin I ntegr Med.Principal component analysis and exploratory factor analysis using SAS software package[J]. Zhong Xi Yi Jie He Xue Bao,Vol.8(2010) No.6,p.591-592.

[6]. Li Yang,Wang Jing,Zhang Yong-jun.Quality analysis method for hot strip rolling based on data mining technology.Chinese Journal of Engineering,Vol.37,Suppl.1:56-64,p.57-58.

[7]. Information on:http://baike.baidu.com/item/FP-growth

[8]. Liu W X,Chen J L,Qu S F,et al.The improved Apriori algorithm.Computer Engineering and Applications.Vol.47(2011) No.11,p.149-151.

[9]. Zhang Y T,Yu Z L,Zhang H X.The frequent item sets miningresearch in association rules.Computer Engineering and Applications.Vol.47(2011) No.3,p.139-141.

[10]. Zhuobo Rong,Dawen Xia,Zili Zhang.Complex Statistical Analysis of Big Data:Implementation and Application of Apriori and FP-Growth Algorithm Based on MapReduce. 2013 IEEE 4th International Conference on Software Engineering and Service Science.Beijing,China.May 23,2013,p.968-970.

[11]. Huang Z Y,Zhu T,Zhang G M,et al.Study on parameters optimization of rough rolling spread model for 1780 mm hot strip.China Metallurgy.Vol.18(2008) No.5,p.24.

[12]. SUN Jian-liang, PENG Yan, LIU Hong-min. Dynamic characteristics of cold rolling mill and strip based on flatness and thickness control in rolling process Journal of Central South University. (2014) 21,p.567-576.

[13]. Sun Y K.Modeling and Control of Hot Strip Mill.Beijing:Metal-lurgical Industry Press.2002,p.52-60.

[14]. Jingbo Yuan,Shunli Ding.Research and Improvement on Association Rule Algorithm based on FP-Growth.2011 4th IEEE International Conference on Computer Science and Information Technology(ICCSIT 2011).Chengdu,China. June 10, 2011, p.391-392.

[15]. Yi Yong Ye. Research and Application of Apriori Algorithm for Mining Association Rules. Advanced Materials Research.Vol(1079-1080).p.737-740.

[16]. Xiaohui Li.Improvement of Apriori Algorithm for Association Rules. Psychological Sciences and Computer Science (PPSCS 2010). Wuhan,China.Novermber 27,2010,p.158-160.