# Complex industrial automation data stream mining algorithm based on random Internet of robotic things

## Lianhe Cui

Published online: 31 Oct 2019.

Submit your article to this journal 

Article views: 396

View related articles 

View Crossmark data 

Citing articles: 1 View citing articles

# Complex industrial automation data stream mining algorithm based on random Internet of robotic things

Lianhe Cui

Department of Engineering, Qiqihar University, Heilongjiang, People's Republic of China

**ABSTRACT**

In recent years, with the continuous development of computer application technology, network technology, data storage technology, and the large amount of investment in information technology, enterprises have accumulated a large amount of data while transforming and improving enterprise management modes and means. How to mine useful data, discover important knowledge and extract useful information has become a hot topic of current research. Industrial big data is significantly different from traditional big data. The traditional big data is based on the Internet environment. Although the data has a high degree of discretization and distribution, its association is relatively simple. The collection of industrial process data is relatively easy, but the mathematical and physical and chemical mechanism models involved make the inherent relationship of data complex, so it is difficult to use common analytical models and methods for processing. In this paper, we propose a complex industrial automation data stream Mining algorithm based on random internet of robotic things, and experimental results show that the proposed algorithm has higher data mining efficiency and robustness.

## 1. Introduction

The development of information technology has triggered the explosive growth of the scale of information production and the speed of communication, making human social life and scientific research enter the era of big data. Studies have shown that the amount of data generated by human activities over the past three years is more than the sum total of the data created by the entire history of human civilization, and this trend is still accelerating [1–4]. In the field of industrial manufacturing, in order to improve the process and control the cost, the conventional method presupposes that the characteristics of the research object are known in advance, and then performs closed-loop control according to the characteristics of the object, so that the output characteristics meet the requirements. Existing manufacturing process modelling methods and automatic control methods are all studied in this way based on a small amount of valuable data. However, many systems in real life are too complicated, and there is no corresponding theoretical knowledge as research support. Their characteristics and behaviours cannot be understood and mastered, and traditional methods cannot play a role. In this case, Jim Gray's data-centric research thinking is equally applicable. For example, in the face of complex industrial production systems, the system's operational process data is preserved by informatizing the complex behaviour of the system.

In the 1960s, in order to meet the requirements of electronic information, information technology began to change from a simple file processing system to an effective database system. In the 1970s, three major models of database systems: research and development of hierarchical, network, and relational databases made significant progress. Since the mid-1980s, the combination of relational database technology and new technologies has become an important symbol of database research and development. In the 1990s, distributed databases became more mature in theory, and distributed database technology was widely used. However, the application of these databases is based on real-time query processing technology. In essence, the query is a passive use of the database, so it still has a long way to go with advanced applications such as analytical forecasting and decision support [5–8]. With the rapid development of information technology, the scale, scope and depth of database applications continue to expand, and data clusters have evolved from single machines to network environments. In recent years, with the rapid growth of data, data analysis tools in existing information management systems have been unable to adapt to new demands. Because the data is

**CONTACT** Lianhe Cui ✉ meikonk7765@163.com 🏛 Qiqihar University, No.42, Wenhua Street, Qiqihar, Heilongjiang, 161006, People's Republic of China

processed in a simple way, the specified data is simply digitally processed, and the intrinsic information contained in the data cannot be extracted. People want to be able to provide a higher level of data analysis, automatically transforming the data to be processed into useful information and knowledge. At present, the emergence of large-scale databases, especially data warehouses, provides an excellent platform for the development and application of data mining technologies. Large-scale databases and data warehouses provide a material basis for the implementation and application of data mining technologies.

The industrial big data analysis platform is a powerful weapon to enhance the competitiveness of enterprises in the industrial field. From the perspective of platform products, data analysis platforms are mostly built on cloud computing and cloud services, with no obvious industrial features. Although China has a large industrial scale, the level of intelligence and information is not high, and the development of industrial big data has just started. According to survey data from the Ministry of Industry and Information Technology, the smart manufacturing readiness rate of China's industrial enterprises is only 5.1%. Compared with developed countries, there is still a big gap in the level of industrialization and informatization, and their industrial big data platform may not be applicable to the current domestic industrial environment. In terms of data accumulation, building a big data platform can optimize the ability to collect, store, mine, and apply large amounts of data in the production process through clustering. This can speed up the processing speed of data, collaboratively manage and analyze the data generated in each link of the production process and improve the utilization of data. Industrial Internet and industrial big data are new concepts put forward by GE in the context of the big data era. Internet-connected sensors collect large amounts of data for analysis, facilitating production and service through a combination of devices, the Internet, and big data. According to General Electric, the sensor has been embedded in 250,000 "smart machines" manufactured by General Electric, including jet engines, power engines, medical equipment and more. The data collected and analyzed by these sensors has great potential for optimizing industrial operations [9–11]. Standard process for data mining is presented in the Figure 1.

(1) Business Understanding: To implement any data mining project, data miners must have an in-depth understanding of the domain knowledge involved in the project. This phase requires extensive research on data mining projects, deep understanding, and then translates this knowledge into tasks that data mining can describe and designs a preliminary plan to achieve the goals.
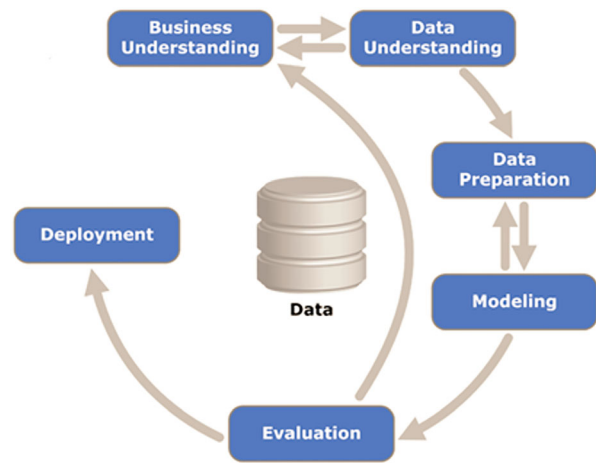


**Figure 1.** Standard process for data mining.

(2) Data Understanding: Data understanding involves understanding the source, shape, and reliability of the data. At the same time, preliminary analysis is needed to understand the preliminary characteristics of the data set. The smooth implementation of the follow-up process relies on the data miner's understanding of the data. The CRISP-DM model divides the data understanding phase into four subtasks: data collection, data description, data exploration and quality verification. The more you master the data, the more targeted the subsequent processes will be.

(3) Data Preprocessing: The task at this stage is to provide an accurate and efficient data set for subsequent data mining models. The specific tasks of data preprocessing are closely related to specific mining tasks and models (algorithms). In general, data cleaning is a necessary step to successfully complete all data mining tasks. Depending on the needs of the algorithm, it is sometimes necessary to standardize and discretize the data. This part of the work is very important, it directly affects the quality of the data being mined.

(4) Modelling: The modelling phase is the narrowly defined data mining phase, which uses data mining algorithms to analyze and process data. The first three processes are all preparing for the modelling phase, and the next two processes are to finish the work of the modelling results. It can be seen that the modelling phase is the core part of data mining. Common data mining models are: clustering model, classification model, rule extraction, time series analysis and so on.

(5) Evaluation: From the point of view of data analysis, a high-quality data mining model has been established at this stage. Before finally extending the model, the model should be evaluated more thoroughly, examining the steps it performs and be confident that it has achieved its goals correctly.

**Table 1.** Three stages of industrial database management.

| | | Manual management | File system management | Database system management |
|---|---|---|---|---|
| Background | Application | Scientific computing | Scientific computing, data management | Large-scale data management |
| | Hardware | No direct storage | Disk | Large capacity disk |
| | Software | No operating system | File system | Database system |
| | Processing method | Batch processing | Real-time processing | Real-time processing, distributed processing, batch processing |
| Features | Data manager | User | File system | Database system |
| | Data sharing | Non-shared | Poor sharing | High sharing and low redundancy |
| | Data independence | Not independent | Poor independence | Highly physically independent and logically independent |
| | Data structure | No structure | Poor structured | Overall structuralization of data |
| | Data control ability | – | – | Database service |
| | Data oriented object | An application | An application | Real world |

(6) Deployment: Apply the data mining model to the actual process and successfully complete the originally set goals.

After data mining was put forward in the 1990s, after more than ten years of research, many new concepts and methods have been produced. Especially in recent years, some basic concepts and methods have gradually become clear, and its research is developing in a more in-depth direction. Three stages of industrial database management is demonstrated in the Table 1.

At this stage, China's intelligent manufacturing has just started, and the existing data storage mode of the enterprise cannot meet the requirements of large-scale data analysis. Therefore, the existing data of the enterprise must be imported into the big data analysis platform through certain methods. In addition, for large-scale manufacturing enterprises, the data generated by enterprises every day is above the GB level. The traditional data import and export method cannot meet the needs of enterprise data import and export, so this paper proposes an industrial data stream mining system based on IORT.

## 2. Random Internet of robotic things

### 2.1. Composition of industrial robot vision guidance system

With the introduction of the visual system, the maturity of the robot technology has led to the use of many industries, and the sensing technology has changed from contact to non-contact. The robotic IoT technology makes it more comfortable to use on the production line. The vision system monitors the product and aggregates the collected information. In addition, visual guidance technology is also widely used in other places. Nowadays, the use of robots in production is increasing, the production line needs to be fully automated, and the flexibility is also increasing. The maturity of visual guidance technology has also led to a continuous reduction in costs, and companies have chosen to use visual systems [12–16].
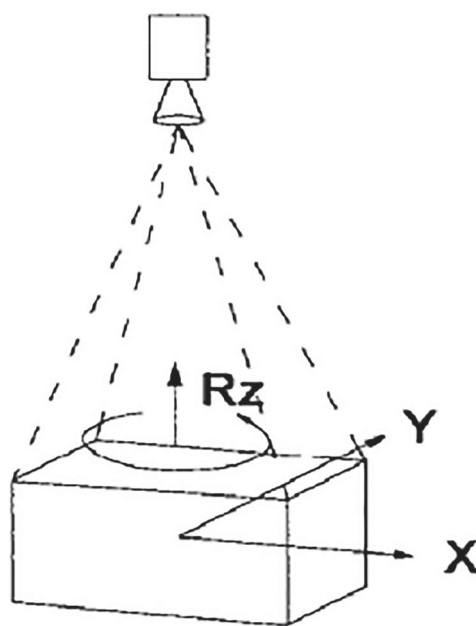


**Figure 2.** 2D vision system schematic.

The 2D vision guidance system is relatively simple. We usually install a camera above the position where the workpiece passes to identify and position the workpiece. The 2D vision system first needs to establish a mathematical model to determine the shape of the different workpieces to calculate the coordinate values of the feature points and as shown in Figure 2, the workpiece is positioned in three degrees of freedom on the X-axis, Y-axis, and Z-axis rotation Rz. Such vision systems are typically used in horizontal transport applications where the workpiece is not positioned, such as belt conveyor. 2D vision system schematic is demonstrated in the Figure 2.

The establishment of the 2.5D vision system model is similar to that of 2D, with only one additional Z-axis height monitoring. However, the acquisition of such information is difficult in visual calculations [17–19]. In some applications, in order to solve this problem, it is usually only a simple addition of the distance sensor, which only measures the deviation of the Z axis. The

2.5D vision system uses a monocular camera to detect changes in the Z-axis height of the workpiece [20–22].

Select the feature points on the workpiece that can easily obtain the distance of the workpiece length. For example, you can select a fixed edge or a hole and select the longer feature on the workpiece as much as possible. This can improve the accuracy. We select 2 lengths of 2 points between L points. Feature points. The following equations are listed according to the principle of similar triangles:

$$\frac{X_1}{L} = \frac{f}{H_0} \tag{1}$$

$$\frac{X_2}{L} = \frac{f}{H_0 + H_1} \tag{2}$$

$$\frac{X_3}{L} = \frac{f}{H_0 + H_1 + H_2} \tag{3}$$

Simplified:

$$X_0 \bullet H_0 = X_2(H_0 + H_1) = X_3(H_0 + H_1 + H_2) \tag{4}$$

Through the above calculations, we can solve the $H_0$ we need. Due to the accuracy error in the motion of the robot, we need to repeat the experiment several times to calculate the $H_0$ with smaller error. After obtaining a more accurate H0, the moving height $Hn$ of the workpiece that passed later can be calculated. The result is as follows:

$$H_n = \frac{X_1 \bullet H_0}{X_n} - H_0 \tag{5}$$

A Cartesian coordinate system is defined in the digital image, where the coordinate $(u, v)$ of each pixel represents the number of columns and the number of rows of the pixel in the image array. $(u, v)$ is a coordinate value of a pixel in pixels in the image digital coordinate system. According to the three-dimensional position of the image point on the image plane, a two-dimensional coordinate system of the image plane expressed in physical units can be established. The x-axis and the y-axis of the coordinate system are parallel to the u-axis and the v-axis, respectively, and the origin is a corner point of the camera optical axis and the image plane [23,24].

In general, the origin is at the centre of the image, but in practice, the origin will have a certain amount of offset. In the two-dimensional coordinate system, we record the coordinates as B [25–27]. In the x-axis and y-axis directions, the physical size of each pixel is the sum $S_x$ and $S_y$. Then the coordinates of any pixel in the image in two coordinate systems can be expressed in homogeneous coordinates and matrix form:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/S_x, 0, u_o \\ 0, 1/S_y, v_o \\ 0, 0, 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6}$$

The inverse relationship of equation (6) is expressed as:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} S_x, 0, -u_o S_x \\ 0, S_y, -v_o S_y \\ 0, 0, 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{7}$$

Therefore, the correspondence between pixel points and world coordinate points is:

$$\begin{cases} X_w = \dfrac{f_x(u - u_o)l \cdot tg\alpha}{f_x(u - u_o) + tg\alpha} \\[2mm] Y_w = \dfrac{-f_y(v - v_o)l \cdot tg\alpha}{f_x(u - u_o) + tg\alpha} \\[2mm] Z_w = \dfrac{f_x(u - u_o)l}{f_x(u - u_o) + tg\alpha} \end{cases} \tag{8}$$

### 2.2. 3d vision guidance system

The 3D system can position the 6 degrees of freedom of the workpiece, so the 3D vision system is the most complex of these types of vision systems, typically used in complex processes and equipment. The 3D vision system is like the two eyes of a person. When positioning the workpiece, it usually takes two cameras with different angles to shoot together. Using only one camera is the goal of the new vision guidance system, which reduces costs [28–30]. The 3D vision system can be used to identify its orientation without touching the workpiece, which is basically suitable for all occasions. There are several main advantages:

(1) Has a wide range of usage. Take the car production workshop as an example, the body-in-white production line can use this vision guidance system. Different production lines have different requirements for the accuracy of the products, as well as the shape and structure of the workpieces. We can choose different numbers of cameras to meet the requirements. For small workpieces, we can use a single camera to achieve spatial positioning. For larger workpieces, such as vehicle positioning, we can use multiple cameras. We can use the collaborative work of multiple cameras to complete the positioning of different parts of the vehicle, thus establishing its spatial position relationship.
(2) Multiple objects can be measured simultaneously. When measuring multiple objects, we can individually measure the individual objects using the reference coordinate system. In this way, the measurements between multiple objects do not affect each other and can better handle the accuracy problem [31,32].
(3) Can be applied in complex environments.
(4) Have the ability to learn. The accumulation of feature templates, the system will remember these as an example, so that the same processing can be used directly later, which makes the recognition rate of the system higher.

(5) The system is highly integrated. The 3D system can connect many robots with colleagues, and its interface is also very diverse.

(6) The system is very scalable. Recalibrate the system when the error is large, without having to teach again. The vision system also automatically saves the template and optimizes it.

## 2.3. Visual system composition of robotic Internet of things

The role of the vision guidance system in the entire production line is mainly to cooperate with the robot to complete the action, it plays the role of connecting the robot and external equipment. In a coherent work, the camera part is responsible for the shooting task, and the information of the shooting is processed to obtain the position information of the current workpiece, and the robot is instructed to perform the correction, and the grasping task is completed after the position is consistent. With the intervention of the vision system, the robot will maintain a posture when grabbing the workpiece, but only need to adjust the position each time to ensure that the grasping task can be successfully completed every time the workpiece position changes [33–35].

The purpose of introducing a visual guidance system is to achieve a high degree of automation of the production line, and to fully utilize the characteristics of high efficiency, flexibility and high level of intelligence. The visual guidance system is a bus system, an identification system, a vision system, a robot system, and an auxiliary system are main components of the vision guidance system. The working principle of the vision system is that the visual part first plays a role, and the photograph of the workpiece is taken when the workpiece is in place, and the coordinates of the feature points selected in advance are calculated after the processing. After that, the correction amount of the workpiece coordinate system is acquired, and finally the correction amount is converted into hexadecimal data recognizable by the robot and the transmission is completed. After receiving the data, the robot will make an action to complete the establishment of the new workpiece coordinate system. In this project, the bus system completes the entire signal transfer task.

(1) Robot system: The main components of the robot system are the robot part, the robot control part, the teaching part and other auxiliary parts. The robot system has two main parts, power and control. Up to eighteen controllers can be installed according to the latest robot control cabinet equipment. In reality, the external axis of the robot generally uses a dedicated motor, and the external axis can make the robot have a lot of freedom, so that the space for the robot to complete the action is larger.

(2) The system of consciousness: the function of the robot lies in the execution, and the part of the vision is the compensation. The camera needs to complete the acquisition task. When the workpiece is in place, the camera collects the photo of the workpiece, extracts the feature points in the photo, and finally obtains the deviation amount, and then transmits the acquired deviation to the robot or to the PLC.

(3) Bus system: In industrial production, the bus acts as a control system tie. If the bus fails during operation, all other devices will be affected and cause downtime. Downtime in industrial production would cause significant losses. The use of various interface technologies and coupling boards can reduce this situation and ensure the normal transmission of data between devices.

## 3. High lights of the proposed methodology

### 3.1. Data stream model and its characteristics

The definition of the data flow model is different for different application environments, and the data flow models under different definitions have their own characteristics and scope of application. Therefore, we need to design a targeted analysis method to achieve various mining tasks.

The data stream can be described as a one-dimensional digital signal, denoted by $S$, in which the signal data arrives in real time, in order, and item by item. $a_t$ represents the data item that arrives at the timestamp $t$, and $F$ is a function that generates the data stream $S$.

$$F : [1, 2, \ldots, t, \ldots] \rightarrow S \qquad (9)$$

Combined with the definition of time series data, the data stream is described as a dynamic, real-time, and unlimited time series. Then the time series data flow model is as follows:

$$F = \{\ldots, a_{i-1}, a_i, a_{i+1}, \ldots\}, a_i = 9(i, I_t) \qquad (10)$$

For monitoring IP addresses, the same IP can access the server multiple times and send multiple packets, because the packet transmission at different times on the same IP link constitutes the Cash Register data stream, as shown below:

$$F = \{\ldots, S_{i-1}, S_i, S_{i+1}, \ldots\}, S_i = S_{i-1} + I_i, a_i = (i, I_i) \qquad (11)$$

The time series data flow model and the Cash Register data flow model have good practical application significance, and they relax the restrictions on the definition of the data flow model. Therefore, these two methods reduce the difficulty of various data flow analysis and mining tasks.

### 3.2. Characteristics of data flow

A data stream is a data model that describes the dynamic characteristics of a data. Compared with a static data set stored on a storage medium, it has the following characteristics:

(1) real-time. The data flow model is a dynamic data generation process, and each data item has a timestamp indicating the order in which the data arrives. The generation, transmission, reception and processing of data are all done in a real-time environment.

(2) Independence. The generation of data streams is only relevant to the associated generation system and is independent of the analytical processing system. The order of arrival of the data stream is independent, and the analysis processing system can only analyze and process the data stream in the order of the timestamp of the data item.

(3) Infinite scale. The continuous and continuous arrival of data streams does not predict the overall data size. The theoretical scale of the data stream can be infinite, and it is impossible to save all the data streams in memory for analysis.

(4) Locality. The characteristics of the data stream determine that the application system cannot predict the overall distribution of the data stream. Therefore, we can only analyze the local features of the data stream and aggregate the low-level local information into high-level local information.

For these characteristics of the data stream, the analysis and processing method of the data stream must make corresponding changes to meet the requirements of the data stream mining task. The design of the data stream mining method needs to follow the following two principles:

(1) Unit time principle. The time that the data stream mining method processes each data item should be fixed, that is, the time complexity of the data stream mining method must be linear, and the processing time must be smaller than the time interval at which the data stream arrives. When the time for processing a single data item is greater than the time when the data stream arrives, it will cause blocking of data reception and processing or loss of data, which will affect the efficiency and accuracy of data processing.

(2) Unit space principle. We need to complete the processing of data streams and the storage and update of intermediate results in a limited memory space. The intermediate result of saving data stream mining and the memory space for processing a single data item should be fixed, that is,

the spatial complexity of the data stream mining method should be constant or linear.

When the intermediate result of data stream mining in memory is large, the data mining update time is longer, and vice versa. Therefore, in the data flow mining process, the balance between the unit time principle and the unit space principle reflects the balance of efficiency and precision of data stream mining. Efficient mining of data streams requires less time to analyze and process the data stream, thereby reducing the size of the intermediate results of the data stream mining in memory to reduce update time. High-precision mining of data streams requires the storage of more detailed, larger-scale intermediate results in memory, which in turn increases the time spent on data mining and the update of intermediate results. Therefore, the final result of data stream mining is an approximate result, and the data stream mining method should be able to balance the efficiency and precision of data stream mining.

### 3.3. Data stream mining method implementation

In the data stream application environment, when the data stream reaches a fast rate and the data size is huge, only using window technology may cause system blocking or data loss, which may reduce the accuracy of data mining. Therefore, sampling technology, as an important data stream processing technology, can extract feature information reflecting the overall data set as a sample data set, and effectively reduce the size of the data set that data mining needs to process. For the characteristics of the locality and scale of the data stream, the size and overall distribution of the data cannot be predicted. Therefore, it is necessary to design a dynamic sampling technique that satisfies the data flow application environment, that is, the sampling process is a dynamic process.

Density-based data stream clustering method can find clusters of the arbitrary shapes, but the algorithm presets too many parameters. The clustering method of data flow based on grid runs very fast, and clusters of any shape can be found, but the clustering quality depends on the selected grid granularity. In order to improve the mining speed of window sliding, rela_table is established in the algorithm CFMoment to store the relationship between frequent non-closed itemsets and frequent closed itemsets. In addition, the algorithm also uses an extended closed-loop enumeration tree based on prefix tree to store data as closed frequent itemsets and related information for w transactions in the stream, further improving mining efficiency and reducing memory consumption.

The reservoir sampling method is a simple random sampling method in which each data stream element is extracted to the sample set with the same probability. The core idea is as follows: maintain a sample

of size *m*, called the "reservoir". By scanning the data stream elements entering the window, the data element is selected into the reservoir with a probability of *m/n*, and the time-stamped data stream in the reservoir is replaced. The reservoir sampling method assumes that the data stream elements are independently and identically distributed, so there is no need to consider the correlation between the data streams, which makes implementation relatively simple.

The reservoir sampling algorithm in the data flow environment is as follows:

*Input: D, Data stream; k, Reservoir sampling scale.*
*Output: S, Sample collection.*
*1: Initialize the reservoir*
    *Inset first k elements from D into S;*
*2: Uniform sampling, update the reservoir*
    *for i = 1,2, . . .*
        *generate a random integer M from the*
            *range [1,i];*
        *if M <= k*
        *insert $d_i$ into S from D;*
        *delete $s_i$ from S;*
        *end*
    *end*

## 3.4. Cardinality estimation: Loglog Counting

First, we need to hash all the elements in the data collection. The hash function here must be guaranteed to be evenly distributed (even if the elements in the collection are not uniform). This premise is the basis of Loglog Counting. The data normally collected by the nonlinear electronic network is continuous in time series and is only active in a small range. When there is interference in the electronic network, the network data transmission will produce a certain packet loss phenomenon, and abnormal data will appear. The abnormal data is data that is different from other data in the electronic network data sequence, and some abnormal data is higher than normal. When the value is one data level, if it is not processed in time, the statistical feature quantity of the nonlinear electronic network model data will be changed, and the number of classifications will be increased.

Using pull up criterion can effectively identify the nonlinear electronic network abnormal data points in the data, but this method is mainly based on network data to describe the whole time series, there are certain error of judgment, and the main reference sliding window to restrict, the identification method of abnormal data in determining if a pull up near point abnormal data points, if it is to change nonlinear electronic network transformer condition, does not carry on the data processing, will use the sliding window average instead of abnormal data. Because other complete item set mining algorithms require more than two scans, and the
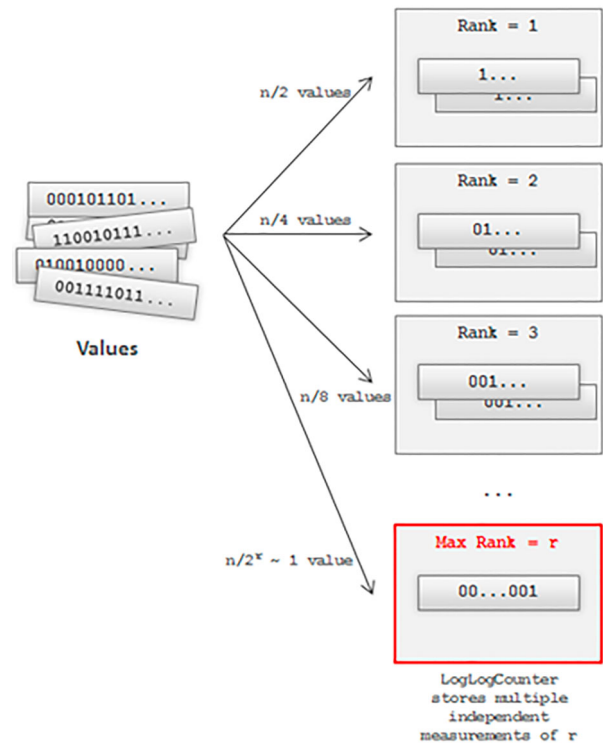


**Figure 3.** Loglog Counting for Data Stream.

base tree needs to be reorganized during the processing of the new batch arrival of the data stream, the calculation time is longer, and the algorithm effectively solves these two problems as a significant improvement in computing time. The algorithm has the following advantages: 1 only need to scan the data set once; 2 the top-down tree traversal process; 3 the maintenance efficiency of the transaction in the sliding window is high; 4 it can accurately find the complete frequent item set.

Under the assumption of uniform distribution, the generated hash value has the distribution ratio in the following figure, because the probability that each bit is 0 or 1 is 1/2, so the more the number of consecutive zeros at the beginning, the smaller the probability of occurrence, the more times you need to try the Bernoulli process. Loglog Counting is based on this principle, based on the maximum number of ranks that appear, to estimate the number of Bernoulli processes (ie, Cardinality). Loglog Counting for Data Stream is demonstrated in the Figure 3.

Suppose we use a two-dimensional hash table, where w is the value space of the hash table, and d is the number of hash functions. For an element, use d hash functions to calculate the corresponding hash value, and increment by 1 on the corresponding bucket. The value of each bucket is called sketch, and then when querying the frequency of an element, we only need to fetch all d sketches, and then take the smallest one as the estimated value. This approach saves data space. w*d is much smaller than the actual number of elements, so there will inevitably be many conflicts. The idea of this method is similar to the bloom filter, which
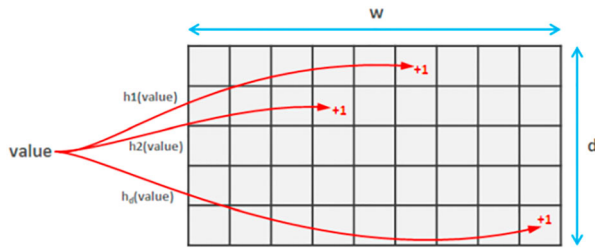
**Figure 4.** Frequency Estimation: Count-Min Sketch.

**Table 2.** Factory resource configuration.

| Component | Quantity |
|---|---|
| Data collection point | 50 |
| Master machine | 1 |
| Task management component | 1 |
| Database management component | 11 |
| Data query component | 2 |
| Data analysis component | 1 |

is to reduce the impact of the punch through multiple hashes. Figure 4 presents the Frequency Estimation: Count-Min Sketch.

## 4. Experiment and analysis

### 4.1. Experiment 1

According to the above estimation of the input parameters of the numerical control equipment factory, we select the resource configuration parameters of each functional component as following Tables 2 and 3:

In this section, bolt tightening data are clustered.

Based on the simulation results, under the premise of factory input parameter estimation, the system bottleneck is first reached as the task management component. If you have more task requests, consider deploying the task management component as a cluster. At this point, the bottleneck point of the system will be determined by the host computer or network bandwidth. Table 4 gives the demonstration.

By analyzing the above table, the cumulative contribution rate of the first four principal factors has exceeded 94% by the fourth principal factor. It can be considered that the first four main factors can represent most of the information of the 14 indicators.

Estimate the application requests that the system needs to process on average every day, such as data query requests, data analysis requests, and so on.

**Table 4.** Comparison of evaluation factors in each hub.

| Ingredient | Total | Initial % | eigenvalue of variance | Accumulated % |
|---|---|---|---|---|
| 1 | 12.153 | 60.76 | | 60.76 |
| 2 | 2.939 | 14.69 | | 75.46 |
| 3 | 2.455 | 12.28 | | 87.73 |
| 4 | 1.287 | 6.46 | | 94.17 |
| 5 | 0.665 | 3.32 | | 97.49 |
| 6 | 0.353 | 1.77 | | 99.26 |
| 7 | 0.079 | 0.39 | | 99.65 |
| 8 | 0.045 | 0.23 | | 99.88 |
| 9 | 0.021 | 0.11 | | 99.98 |
| 10 | 0.004 | 0.20 | | 100.00 |

**Table 5.** System parameters estimation diagram.

| Parameters | Meaning | Value | Unit |
|---|---|---|---|
| WriteQRate | Message middleware queue speed | 10 | M/second |
| MOMworkingrate | Message middleware handles message speed | 30000 | M/second |
| MOMthroughtput | Message middleware throughput | 100 | M/second |
| SAworkingRate | Master control request rate | 3000 | kb/second |
| SearCOMworkingRate | Data query component request rate processed by a single node | 3000 | kb/second |
| StaCOMWorkingRate | Data analysis component single node processing request rate | 3000 | kb/second |
| TMworkingRate | Task management component single node processing request rate | 300 | 1/second |
| dataCOMworkingRate | Database management component single node processing request rate | 300 | kb/second |
| LANBandwidth | LAN bandwidth | 100 | M/second |

According to the statistical information of the CNC machine tool factory, the average demand for data query and data analysis is about 20 times per day, and the message size of a single request is only between tens of bytes and 1 kb.

The other parameters of the simulation process are shown in Table 5:

### 4.2. Experiment 2

In the experiment we match the template and compare the coordinates of the workpiece image. Map the image coordinates to the robot's coordinate system to get the coordinates of the workpiece on it. The Kalman filter is used to estimate the position of the workpiece in the robot's gripping area. After several tests, the calculated data is compared with the position data of the

**Table 3.** Analysis results.

| Category | Proposed | Stat = 0 | Stat = 1 | STKmeans | Stat = 0 | Stat = 1 |
|---|---|---|---|---|---|---|
| 0 | 54315 | 54311 | 4 | 153835 | 153814 | 21 |
| 1 | 64542 | 64540 | 2 | 90 | 14 | 76 |
| 2 | 25485 | 25485 | 0 | 291 | 291 | 0 |
| 3 | 29441 | 29425 | 16 | 24 | 24 | 0 |
| 4 | 68275 | 68275 | 0 | 40 | 40 | 0 |
| 5 | 58101 | 58101 | 0 | 164 | 163 | 1 |
| 6 | 545 | 468 | 77 | 146260 | 146295 | 1 |
| Total | 300704 | 300605 | 99 | 300704 | 300605 | 99 |

**Table 6.** Workpiece positioning and grabbing experiment results.

| Serial number | Visual positioning experimental results | End position of the robot |
|---|---|---|
| 1 | (420.25, 136.97, −25.08) | (419.9, 135.7, −25.1) |
| 2 | (410.36, 190.53, −25.16) | (410.7, 191.5, −24.9) |
| 3 | (460.36, 201.35, −25.33) | (469.1, 201.5, −26.6) |
| 4 | (465.35, 260.32, −25.43) | (464.9, 259.1, −25.5) |
| 5 | (532.13, 190.77, −25.51) | (531.4, 199.3, −25.1) |

TCP movement point touched by the robot end within 1 mm. Table 6 gives the simulation.

## 5. Conclusion

With the continuous development of industrial informatization and the explosive growth of industrial big data, the demand for decision-making of data support has been continuously improved, and industrial big data analysis has become a very important subject. In this paper, the application of machine vision in industrial robots is deeply studied, and its principle and implementation method are studied. This paper focuses on the construction of enterprise data flow analysis platform and relies on the historical data of an enterprise to introduce the general flow of data analysis for big data platform. In order to improve the effectiveness of the data stream mining method in the industry, this paper proposes a complex industrial automation data stream mining algorithm based on random internet of robotic things. Based on the principle of industrial camera imaging, we analyze the method of solving internal and external parameters of industrial cameras and study the method of hand-eye calibration in industrial robot vision system. According to the IORT technology and the computer vision method, we establish a relationship model between the coordinate systems in the visual guidance system and determine the positional relationship between them. Experimental results show that the proposed method has higher efficiency and robustness.

## Disclosure statement

No potential conflict of interest was reported by the author.

## References

[1] Jiao J, Gao W, Han Y. (2018). The nearest neighbor information estimator is adaptively near minimax rate-optimal. In Advances in Neural Information Processing Systems. p. 3160–3171.

[2] Cuzzocrea A, Gaber M M, Shiddiqi A M. Adaptive data stream mining for wireless sensor. Proceedings of the 18th International Database Engineering&Applications Symposium. ACM, 2014. p. 284–287.

[3] Aston N, Liddle J, Hu W. Twitter sentiment in data streams with perceptron. J Comput Commnications. 2014;20(3):11.

[4] Allen TT, Sui Z, Akbari K. Exploratory text data analysis for quality Hypothesis generation. Quality Engineering. 2018;30(4):701–712.

[5] Silberschatz A, Korth HF, Sudarshan S. Database system concepts. 3rd ed. New York: McGraw-Hill; 1997.

[6] Ramakrishnan R, Gehrke J. Database management systems. 2nd ed. New York: McGraw-Hill; 2000.

[7] Berthold M, Hand DJ. Intelligent data analysis: an introduction. *Springer Verlag*. 1999.

[8] Tongming L. Data mining technology and its application. Beijing: National Defense Industry Press; 2001.

[9] Lee R, Luo T, Huai Y, et al. Ysmart: Yet another sql-to-Map/Reduce translator//Distributed Computing Systems (ICDCS), 2011 31st International Conference on. IEEE, 2011, p. 25–36.

[10] Feng J. Research and application of massive data processing based on Hadoop platform. Beijing: Beijing University of Posts and Telecommunications; 2012.

[11] Bing W. Application of large-scale process industry group-level production data platform. Sci Technol Enterprises. 2014;1:129–130.

[12] Qiu Y, Bi Y, Li Y, et al. High resolution remote sensing image denoising algorithm based on sparse representation and adaptive dictionary learning. In: Computational vision and Bio Inspired computing. Cham: Springer; 2018. p. 892–901.

[13] Wang H, Li Z, Li Y, et al. Visual saliency guided complex image retrieval. Pattern Recognit Lett. 2018. doi:10.1016/j.patrec.2018.08.010.

[14] Parker NN, Allen TT, Sui Z. K-means subject matter expert refined topic model methodology (No. TRAC-M-TR-17-008). TRAC-Monterey Monterey United States, 2017.

[15] Allen TT, Sui Z, Parker NL. Timely decision analysis enabled by efficient social media modeling. Decis Anal. 2017;14(4):250–260.

[16] Sui Z. Hierarchical text topic modeling with applications in social media-enabled cyber maintenance decision analysis and quality hypothesis generation [Doctoral dissertation]. The Ohio State University; 2017.

[17] Gupta BB, Yamaguchi S, Agrawal DP. Advances in security and privacy of multimedia big data in mobile and cloud computing. Multimed Tools Appl. 2018;77(7):9203–9208.

[18] Baran R, Partila P, Wilk R. (2018). January. Automated text detection and character recognition in natural scenes based on local image features and contour processing techniques. In International Conference on Intelligent Human Systems Integration, p. 42–48. Springer, Cham.

[19] Zhang C, Liu J, Chen F, et al. Dependency-and similarity-aware caching for HTTP adaptive streaming. Multimed Tools Appl. 2018;77(1):1453–1474.

[20] Niu Z, Hua G, Wang L, et al. Knowledge-based topic model for unsupervised object discovery and localization. IEEE Trans Image Process. 2018;27(1):50–63.

[21] Zhang S, Wang H, Huang W, et al. Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG. Optik. 2018;157:866–872.

[22] Wang N, Gao X, Tao D, et al. Facial feature point detection: A comprehensive survey. Neurocomputing. 2018;275:50–65.

[23] Chen Q, Zhang G, Yang X, et al. Single image shadow detection and removal based on feature fusion and multiple dictionary learning. Multimed Tools Appl. 2018;77(14):18601–18624.

[24] Hui Z, Wang X, Gao X. (2018). Fast and accurate single image super-resolution via information distillation network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 723–731.

[25] Han Y, Jiao J, Weissman T. (2018). Local moment matching: a unified methodology for symmetric functional estimation and distribution estimation under Wasserstein distance. arXiv preprint arXiv:1802.08405.

[26] Li X, Feng Z, Zai J, et al. Incorporation of Co into MoS2/graphene nanocomposites: one effective way to enhance the cycling stability of Li/Na storage. J Power Sources. 2018;373:103–109.

[27] Singh AK. Robust and distortion control dual watermarking in LWT domain using DCT and error correction code for color medical image. Multimed Tools Appl. 2019: 1–11. doi:10.1007/s11042-018-7115-x.

[28] Kuznetsova I, Lugmayr A, Holzinger A. Visualisation methods of hierarchical Biological data: A survey and Review. Int SER Inform Syst Manage Creative eMedia (CreMedia). 2018;2017(2):32–39.

[29] Jiao J, Han Y, Weissman T. Minimax estimation of the $ L\_ \{1\} $ distance. IEEE Trans Inf Theory. 2018;64(10):6672–6706.

[30] Shamsolmoali P, Jain DK, Zareapoor M, et al. High-dimensional multimedia classification using deep CNN and extended residual units. Multimed Tools Appl. 2019;78(17):23867–23882.

[31] Jiao J, Gao W, Han Y. (2018). The nearest neighbor information estimator is adaptively near minimax rate-optimal. In Advances in neural information processing systems, p. 3156–3167

[32] Aljawarneh S, Yassein MB. A multithreaded programming approach for multimedia big data: encryption system. Multimed Tools Appl. 2018;77(9):10997–11016.

[33] Alvarez-Ramirez J, Santibanez V, Campa R. Stability of robot manipulators under saturated PIDcompensation. IEEE Trans Control Syst Technol. 2008;16(6):1333–1341.

[34] Kang HJ, Ro YS. Robot manipulator modeling in Matlab-SimMechanics with PD control and onlinegravity compensation. International Forum on Strategic Technology. October 13, 2010, Ulsan, Korea, p.446–449.

[35] Su Y, Parra-Vega V. Global asymptotic saturated output feedback control of robot manipulators. Proceedings of the 7th World Congress on Intelligent Control and Automation; 2008 June 25–28; Chongqing, China. p. 3445–3450.