2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)

Cigarette factory big data collection and cleaning

First author: Wenchao Huang[1]
China Tobacco Shandong Industrial Co., LTD Qingdao cigarette factory
Qingdao,China
472104959@qq.com

Co-First author: Hongyu Yuan[1]
Shandong Industrial Internet Development Research Center
Qingdao,China
* Corresponding author:hongyumsn@163.com

Xuechao Tang[2]
China Tobacco Shandong Industrial Co., LTD Qingdao cigarette factory
Qingdao,China
15712776407@163.com

Peichen Wang[3]
China Tobacco Shandong Industrial Co., LTD Qingdao cigarette factory
Qingdao,China
1287129602@qq.com

Wenkun Zhao [4]
China Tobacco Shandong Industrial Co., LTD Qingdao cigarette factory
Qingdao,China
2838925485@qq.com

Mengting Zhao[5]
China Tobacco Shandong Industrial Co., LTD Qingdao cigarette factory
Qingdao,China
emailzmt@163.com

***Abstract*—Against the background of the current rapid development of industrial Internet technology, the integration and application of new-generation information technology is penetrating the cigarette manufacturing industry more and more closely. This study focuses on the big data acquisition and cleaning aspects of the production and manufacturing process in cigarette factories, aiming to collect production quality, and equipment data in real-time from diversified data sources, such as sensors, equipment logs, and quality inspection systems, by means of efficient data capture and preprocessing techniques. The study employs advanced data capture methods to ensure real-time and accurate data acquisition. Further, the quality and usability of the data were significantly enhanced by meticulous data cleaning steps, including de-weighting, error correction, and data completion. Clean data acquisition has a significant impact on optimizing production decisions, enhancing quality control, and accurate equipment maintenance in cigarette factories. With the advancement of data networking, cloud storage, and platform integration, the construction of a big data platform is particularly important, which provides a solid foundation for digital transformation and efficiency improvement in the cigarette manufacturing industry.**

Keywords-industrial internet; big data; data collection; data cleaning

I. INTRODUCTION

In the traditional cigarette factory production model, the role of data is often neglected, limited by the ability to collect, store, transmit, and analyze, and only as a supplement to the business process. However, with the rapid development of information technology, the enhancement of algorithms and arithmetic power enables us to effectively identify, measure, and manage the huge amount of data in the production process, thus tapping into the huge potential value of data. Data has risen to become a key production factor in the cigarette industry. In cigarette factories, the application of big data collection and cleaning technology provides strong support for enterprise digital transformation [1]. By building a digital management platform, enterprises can realize the interconnection of R&D, production, management, and service links, and promote the extensive collection of data from equipment, workshops, logistics, and so on. This process builds an effective connection between the whole life cycle, the whole elements, the whole industrial chain, and the whole value chain, forming a data flow loop [2]. The loop helps production management realize intelligent decision-making through state perception, real-time analysis, scientific decision-making, and precise execution, and significantly improves risk perception, prediction, and prevention capabilities. Compared with the drawbacks of the traditional management mode that relies on personal experience and intuition and poor information sharing, digital management takes business digitization as the starting point, combines with advanced data collection and cleaning technology, realizes centralized control of employees and business and optimal allocation of resources, and effectively improves the enterprise's key resource management capabilities. This study focuses on big data acquisition and cleaning technology in cigarette factories, aiming to provide data support for enterprises and help their innovative development in the era of the digital economy.

II. INDUSTRIAL BIG DATA ACQUISITION TECHNOLOGY

A. *Data Acquisition System Architecture*

The application of industrial big data acquisition technology in the cigarette manufacturing industry involves the use of advanced sensors, network communication means and data processing techniques to capture, transmit, and save massive data information from the cigarette production line in real-time or on a regular basis. These data cover a wide range of aspects, such as machine operation data, production process data, cigarette quality data, environmental variables, etc., and constitute the core elements for cigarette factories to realize the industrial Internet and intelligent manufacturing. Through these technologies, cigarette factories are able to monitor the production status more accurately and improve the manufacturing efficiency and quality control level [3].

979-8-3315-0976-7/25/$31.00 ©2025 IEEE

B. *Data Acquisition Technology*

1) Conventional Data Acquisition Technology

It mainly includes the SCADA system and OPC UA.

In cigarette factories, SCADA (Supervisory Control and Data Acquisition) systems are widely used to collect and control various production equipment (PLC devices such as S7-300, S7-400, logix5500, etc.) to realize real-time monitoring of the production process. This system can effectively monitor the key quality indicator parameters on the production line to ensure the stability of the production process and product quality [4]. However, the SCADA system also faces some limitations in its application:

a) Extensibility Limitations

Expansion of monitoring points: when new monitoring points need to be added to cover more production areas or equipment, the SCADA system may not be able to be easily expanded. This usually involves complex network configurations, additional hardware purchases and installations, and corresponding adjustments to the system software, processes that are not only time-consuming but also costly.

Device Integration: With the popularity and application of smart sensor devices, SCADA systems may encounter difficulties in integrating these devices. Intelligent sensors can provide richer data, but the SCADA system may lack compatibility or sufficient processing power to make full use of this data, resulting in the system can not maximize the role of intelligent sensors.

Visual Application Requirements: The development of smart factories is becoming more and more demanding, especially in terms of visual surveillance. Algorithmic analysis based on video models, such as machine vision inspection, product defect identification, etc., requires real-time access to video streams and complex data processing. SCADA systems have limitations in realizing these functions because they are usually not designed to process large-scale video data or perform advanced image analysis.

Real-time monitoring challenges: Operators need to view production conditions in real time through terminals in order to respond quickly to any problems on the production line. However, SCADA systems may not have sufficient processing power or user interface flexibility to support the access and analysis of such real-time video streams, limiting the efficiency of operators in monitoring the production process.

b) Insufficient flexibility:

Rigid system architecture: SCADA systems are usually designed with a fixed architecture, which makes the process of adjusting and upgrading the system complicated and time-consuming when the production process changes or when new production equipment needs to be integrated. For example, if you want to change the layout of the production line or add new production links, you may need to reprogram and configure the entire SCADA system.

Hardware dependency: SCADA systems often rely on specific hardware devices, such as specific PLCs or RTUs (Remote Terminal Units). When these hardware devices need to be replaced or upgraded, the SCADA system may not be directly compatible with the new hardware, requiring additional adapters or software upgrades, which increases system maintenance costs and downtime.

High degree of software customization: SCADA system software is usually highly customized to meet specific production requirements, which means that any small changes may require professional software developers to adjust, which not only increases operating costs, but also extends the problem-solving cycle.

Integration Difficulties: SCADA systems may encounter difficulties when integrating with third-party systems or emerging technologies (e.g., IoT devices, advanced data analytics tools, etc.). These integration challenges limit the interoperability and data utilization efficiency of SCADA systems in a smart factory environment.

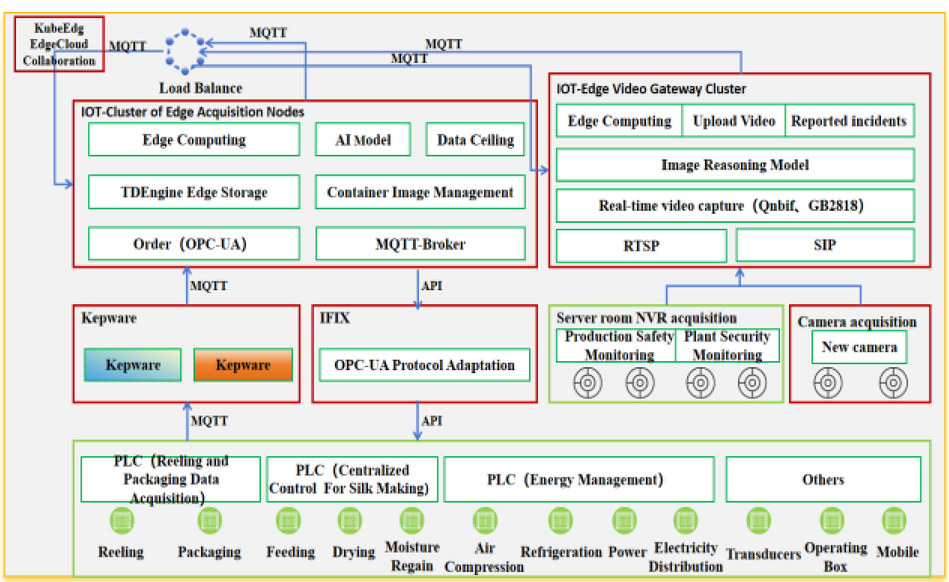

Figure 1. IoT system architecture diagram

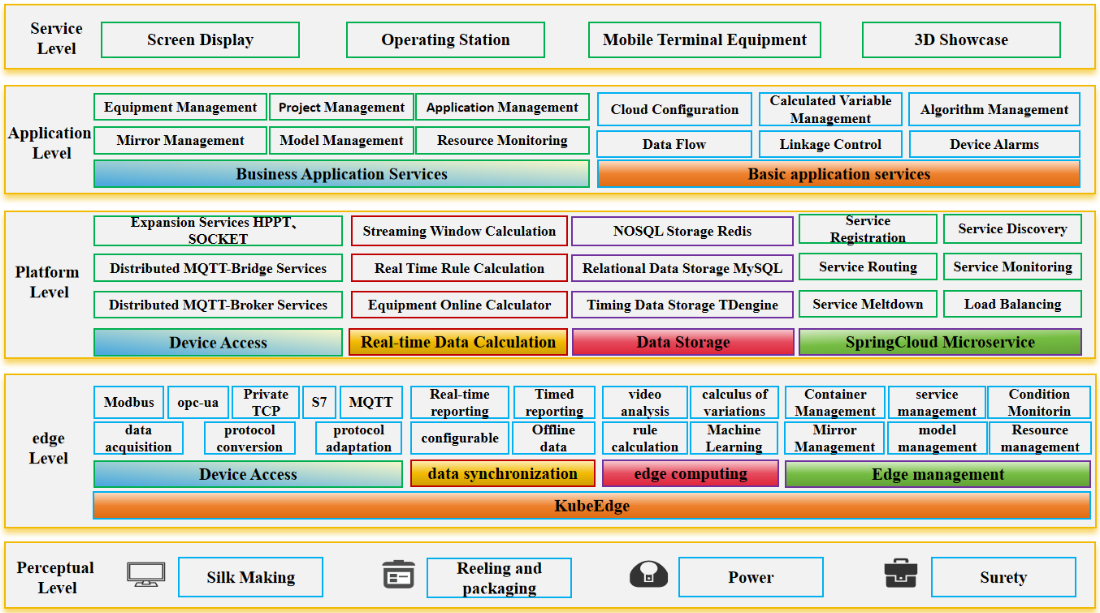

Figure 2. Functional modules of IoT platform

2) IoT data acquisition technology

a) Concept

IoT components are positioned in the middle core layer of IoT technology, and their main role is to connect intelligent devices downward and take up the application layer upward [5]. Growing with data as nutrients and empowering data to downstream applications presents a logic of gradual ascension of data value from upstream terminals to downstream users. Parallel processing of millions of data per second, based on a new pipeline data processing model, realizes data flow as a service, which is a wise choice for enterprise digitalization and intelligent transformation. The IoT component platform is based on a cloud-native environment, supports various connection protocols, and supports data collection and reverse control functions on the edge host, providing functions and intelligent services for edge device applications [6].As shown in Figure 1.

b) Component Functions

IoT contains functions such as terminal management, collection protocols, cluster management, and message queues [7].As shown in Figure 2.

c) Design Goals

- Rich device access support

Cover mainstream protocols and support customized protocols. Support mainstream Modbus TCP, Modbus RTU, OPC UA, MQTT, RTSP, and customized TCP protocols.

- Ultra-large-scale data access, and high-performance processing capabilities

Support for millions of equipment measurement point access, the whole link nodes support cluster deployment, with dynamic capacity expansion, load balancing, and asynchronous processing capabilities.

- Streaming computing capability with SQL support

The platform supports SQL to realize the calculation of the data reported by the equipment and supports the sliding time window.

- Powerful rule engine expression support

The platform supports operators to customize expressions for device alarms, linkage control, and data flow. The rule engine has a rich set of built-in functions and extended custom functions, which is convenient to use out of the box.

- All-round deployment of online status and display

Using the configuration application developer and the map application developer, through simple WYSIWYG configuration or drag-and-drop LowCode/NoCode development mode, quickly realizes the acquisition, monitoring, alarm, analysis, and other functions.

- Deep integration of artificial intelligence

Support for custom functions, artificial intelligence models, real-time predictive reasoning based on mechanistic models and artificial intelligence, support for hundreds of machine learning operators, as well as various types of custom operators.

- Open and Security Capabilities

Provide 400+ open APIs to facilitate integration, sharing, and secondary development by customers or partners. Device access key authentication, whole process data encryption, API access authorization authentication, and other data security.

- Cloud-side synergy

Based on K8S&KubeEdge, it realizes cloud-edge collaboration, extends computing power to the edge, and supports device access, streaming data processing, video analysis, voice image recognition, and other capabilities at the edge.

d) Internet of Things Data Access

The IoT data access system is shown in Figure 3.

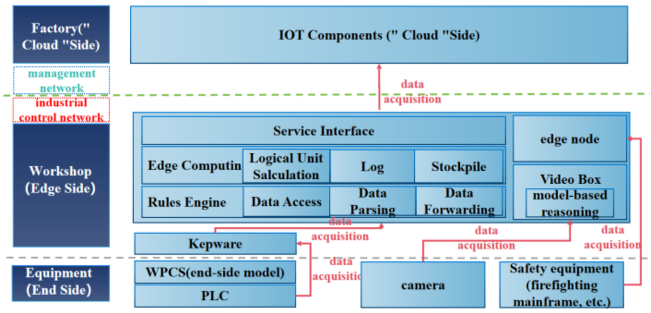

Figure 3. Diagram of IoT data access system

- Real-time data access

Unified access and reverse control of each sensing device and PLC device of the tobacco factory's silk-making, rolling package, and power workshop are carried out through edge computing. The collected data will be uniformly reported to the IOT cloud through MQTT, and then the server unified management system will provide external services through various types of external service interfaces. Among them, the IOT edge end supports Modbus (water, electricity, gas, steam, and other sensors), OPC UA (power shop data system), Siemens S7 (production equipment controller), custom TCP, and other mainstream industrial protocols.

- Video streaming access

Realize real-time monitoring and video analysis and early warning of the internal cameras of the tobacco factory (production process material monitoring cameras, plant area placement monitoring cameras). Edge intelligent video gateway supports GB28181, Onvif, RTSP, RTMP, and other video-related transmission and control protocols. Currently, 89 monitoring cameras are connected to the tobacco storage cabinet in the silk workshop, which makes it convenient for operators to view the status of the materials in the cabinet during the production process, and with the video algorithm, abnormalities of the materials in the cabinet are found in a timely manner.

- Data Push Application Platform

The platform integrates three communication protocols, namely MQTT, Kafka and HTTP, to meet different communication requirements.The MQTT protocol, with its lightweight and low power consumption, is very suitable for transmitting data with high real-time requirements, such as equipment sensor information, production task instructions, etc. The Kafka protocol performs well in processing large-volume data, ensuring high reliability and no data loss, and it is It is mainly used for processing production batch information, equipment operation status records, product quality data, etc., which provides a solid foundation for data analysis and subsequent processing. As for the HTTP protocol, due to its simplicity and ease of use, it is used to handle non-real-time operations, such as configuration update of equipment parameters and process flow, uploading of production reports and quality analysis reports, as well as querying of key information, such as production plan and material inventory. This combination of communication protocols not only ensures real-time transmission and efficient processing of production data, but also enhances the flexibility of the entire platform's configuration and the convenience of information query.

3) Data Center Data Integration Services

The data integration service includes two parts: real-time data acquisition and results-based business data acquisition. Among them, real-time data acquisition can support PLC, OPC UA, MQTT, and other mainstream application layer protocols for data acquisition; the business system can be extracted, converted, and loaded into the database through ETL, which for the business data with large data volume, message queuing + streaming data processing is used to complete the way [8]. Comprehensive access to all types of business data in ERP, MES, TsPM, and other systems. It provides the functions of regular batch extraction and real-time collection, opens up the data channel of the whole plant, completes the data governance and integration of the access system, and provides data-sharing services for each business system of the whole plant. As the underlying basic support service, it is the basic component of the data platform. By providing multi-source data integration capability, various types of structured and unstructured data are converged and accessed in a unified way, so as to build an efficient, easy-to-use, continuously stable, and scalable data transmission channel for the construction of the data platform [9].As shown in Figure 4.

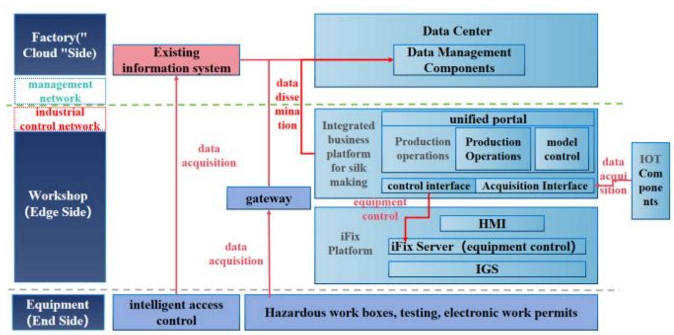

Figure 4. Data center data integration services

For large amounts of data and high requirements for timeliness, use message queuing + streaming data processing to complete data cleaning and checking, standardization, modeling, abstraction, aggregation, etc. For non-real-time data with a large amount of business data, this method can also be used to complete data processing and improve data processing efficiency [10].

a) Data source management module

The data source management module is mainly responsible for maintaining the basic information of various data components supported by the platform, including the addition, configuration, and management of data components. This involves data components including but not limited to relational databases (e.g. PostgreSQL, GreenPlum, Oracle, MySQL) and plug-ins such as Kafka, covering key configuration information such as database IPs, ports, and access rights. The module is used to access relevant data sources such as the cigarette factory logistics system, quality data, MES silk production data, 5K intelligent manufacturing data, intelligent edge job management, the TSPM-management library, Wenlong-real-time data, and OneCard-Security Access Control.

b) Synchronization Task

The Synchronization Task module provides a task list where data synchronization tasks can be created, saved, edited, and scheduled to achieve real-time data synchronization.

c) Integration Tasks

Integration tasks build data integration processes through drag-and-drop operations, including nodes for integration sources, preprocessing, and integration targets. The integration task acquires information such as cigarette factory personnel shift-ATS, special inspection of filament quality, sensory evaluation, equipment protection management, process standard version, business quantitative data, etc. to support the upper layer application of the data platform.

III. INDUSTRIAL BIG DATA CLEANING TECHNOLOGY

A. Data Cleaning Concept

Data cleaning refers to the cleaning, calibration, conversion, and integration of collected data to improve data quality and provide a reliable database for subsequent data analysis. The data cleaning needs involved in the cigarette factory are mainly the processing of real-time data, process data validity marking, abnormal data processing, and missing data processing.

B. Data Cleaning Methods

1) Data cleaning based on the Internet of Things (IoT)

The Internet of Things provides IT and OT data collection capabilities of multiple protocols, especially for the collection of point data represented by time series, and at the same time provides diversified data cleaning means, combined with visual data cleaning capabilities, can be a one-stop solution to data cleaning and development issues.

The platform is divided into 4 types of templates, which are linkage control, flow calculation, data flow, and alarm configuration. The data cleaning function nodes include JSON message splitting node, JSON attribute modification node, data alignment node, judgment branching node, Avro format conversion node, JSON data parsing node, updating variable node, and equipment counter-control (Rest API) node.

a) Methodology

In this paper, Standard Deviation, Mean Value, Deviation, Absolute Deviation, Absolute value of standard deviation, Standard deviation judgment, Maximum value, Minimum value, Median to clean the data, and the calculation formula is shown in (1)-(5).

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - u)^2} \tag{1}$$

$$\mathrm{u} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

$$\Delta = \mathrm{u} - \text{standard value} \tag{3}$$

$$|\Delta| = \mathrm{u} - \text{standard value} \tag{4}$$

$$\Delta\sigma = \sigma - \text{standard value} \tag{5}$$

b) Indicators

The indicators involved in IoT-based data cleansing are shown in Table 1.

TABLE I. Real-time data cleansing

Parameter 1	Parameter 2	dependency condition
Inlet moisture	outlet temperature	1. MES determination rules (batch start time, batch end time) 2. Equipment downtime; restart time
Smoke flake flow rate	Cumulative amount of tobacco flakes	
Instantaneous accuracy	Accuracy cumulative value	
furnace wall temperature	Hot Air Velocity	
head pressure	Mixed Silk Moisture	
Filling flow rate	Cumulative volume of water added	
Ratio of material to liquid	Cumulative volume of material and liquid	
hot air temperature	Export moisture	
vapor flow	vapor pressure	

2) Data Cleaning Based on Data Platforms

The data processing platform can flexibly and conveniently configure tasks, support diversified data processing capability components, have the ability to schedule full-process visualization jobs, and have a scheduling engine combined with reliable, stable and high-performance distributed scheduling management components such as Yarn, to flexibly and quickly access the data for processing. Cleaning component management is mainly for the data cleaning process of preset or custom components for the management of DIPE task configuration, which includes but is not limited to format conversion components, content conversion components, increase column components, dictionary conversion components, connection components, and so on [11].

a) Real-time task management

- DIPE real-time task optimization

In the cigarette factory data cleaning project, the DIPE real-time task builds an efficient data processing flow by dragging and dropping the rule components and fills in the corresponding configuration parameters of the cigarette process data cleaning rules in order to create a streaming Java computing task to realize the instant cleaning and processing of cigarette production data. The specific process architecture is modeled as follows.

$$D_{final} = \mathrm{Load}(D_{join}) \tag{6}$$

$$D_{join} = \{(T(d_1), T(d_2)) | d_1 \in D_{clean1}, d_2 \in D_{clean2}, d_1[K] \in D_1[K]\} \tag{7}$$

$$D_{clean1}, D_{clean2} = \{C(d_i) | d_i \in D_{extract1}, D_{extract1}\} \tag{8}$$

$$D_{extract1}, D_{extract1} = \{d_i \in D | P(d_i)\} \tag{9}$$

Where D_i is the original dataset, $D_{extract}$ is the subset extracted in D_i, P is the conditional function for extraction,

D_{clean} is the removal of invalid, incorrect, or incomplete data A is the removal of invalid, erroneous or incomplete data from $D_{extract}$, C is the cleaning function, $D_{transform}$ is the conversion to another form of data, T is the transformation function, D_{merge} is the merged dataset, M_i is the merge function, D_{join} is the dataset connected based on certain keys, K_i is the association key, and D_{final} is the dataset written to the target data store.

- Flink real-time task deployment

Write FlinkJar and perform FlinkJar remote scheduling execution through the scheduling engine. Configure the basic attribute information of the task, upload and write the Jar package containing the cleaning rules of cigarette production data, and carry out the cleaning task scheduling and processing.The specific process architecture is modeled as follows.

$$\text{Result} = S(\text{write}(\acute{D}), R, T) \tag{10}$$

$$\acute{D} = \{f(d_i) | d_i \in D, C(d_i)\} \tag{11}$$

Where write is the write function, S is the scheduler, R is rule, T is time, $\acute{D}$, C is regular function.

b) Offline task management optimization

- Timer Management

Realize the timer to add, edit, delete, enable, disable, and add triggering offline jobs for timers for offline task timing scheduling.

- DIPE offline task deployment

In the data cleaning project of the cigarette factory, the DIPE offline task realizes a series of processing processes such as data extraction, cleaning, conversion, merging, association, and loading by dragging and dropping rule components, drawing data processing flow, and filling in rule configuration parameters.

- SQL stored procedure

In the cigarette factory data cleaning project, the list of SQL stored procedures supports script query, add, batch submit, single submit, withdraw, delete, and other operations. It can be filtered according to conditions such as the script name, creation time, script status, script type, and so on. The list information contains the script name, creation time, script type, invocation, script status, etc., and the operations include submission, withdrawal, editing, viewing, deletion, and so on.

IV. Application Cases of Industrial Big Data Collection and Cleaning

A. State awareness of total factors of production

In the silk-making workshop of the cigarette factory, the in-depth application of the state awareness of total factors of production has built an all-round and three-dimensional intelligent monitoring network. This network is like the brain and nervous system of the silk workshop, capturing in real time the operating status of each silk production line equipment, the flow trajectory of each batch of tobacco and tobacco materials, the details of each change in the processing environment, as well as the consumption of energy by each point of the tobacco machine equipment. Through the integration, analysis, and intelligent processing of these massive data, the production process of the silk-making workshop becomes transparent and refined, and the managers are able to accurately grasp every detail of the production site as if they have a perspective eye, discover potential faults and hidden dangers in time, prevent production interruptions, and ensure a smooth and efficient production process.

This kind of all-factor state awareness not only improves production efficiency, but also plays an important role in guaranteeing product quality, improving energy utilization efficiency, strengthening safe production, and environmental protection. It makes the silk workshop change from traditional passive management to active and preventive management, injects a strong impetus for the intelligent upgrading and digital transformation of the workshop, and also provides strong support for the enterprise to maintain its leading position in the fierce market competition.

B. Fully automated production batch coordination

In the modernized cigarette factory silk workshop, the application of the fully automatic batch coordination system has completely changed the traditional production mode and pushed the production process to a whole new level. Through the integration of advanced information technology and automation equipment, the system realizes the whole process of automation management from the raw material formula library to the silk production line to the silk finished product storage. In this system, each batch of tobacco grade production can realize precise synergy with other batches, whether it is the proportion of raw material formula library, the process parameters of the silk production process, or the operation status of the equipment, can be monitored in real-time and intelligently adjusted.

The fully automated batch coordination system collects production data in real-time by integrating intelligent sensors, the silk production line host PLC, QR codes of the tobacco box, and camera video surveillance, and realizes data integration and interconnection by using communication protocols, such as industrial Ethernet, etc. The control system monitors and analyzes data in real-time and evaluates the stability of the production through tools such as SPC, etc., and at the same time, it defines and traces each batch and adjusts the batch automatically. The system also defines and tracks each batch, automatically adjusts the order of operation between batches and production lines, realizes process coordination, provides decision support for exception handling and predictive maintenance, optimizes the production process through closed-loop control and continuous improvement mechanism, and ensures the smooth operation of the production process.

C. Prediction of hidden dangers in the whole production guarantee

The application of big data collection and cleaning technology is closely connected with the establishment of the static diagnostic model. Combined with the analysis of the performance parameters of the host equipment and the historical data of the key parameters, the big data acquisition is responsible for collecting various data during the operation of the equipment, including temperature, pressure, and other key indicators. Through pattern recognition technology, this data is used to build a static diagnostic model, which is capable of automatically detecting potential problems, such as temperature anomalies, pressure anomalies, and so on.

Data cleaning plays a vital role in this process, ensuring that the data fed into the model is accurate and reliable. The cleaned data is passed through set detection rules and criteria, enabling the system to quickly identify potential problems. Once an anomaly is detected, the system automatically triggers an alarm and, in conjunction with digital twin technology, displays information about the type and location of the anomaly, helping operators to take action quickly.

During the production process, big data collection and cleansing of real-time monitoring equipment status data, combined with equipment fault diagnosis and predictive maintenance judgment, can provide early warning of possible future abnormalities. This early warning mechanism enables operators to take preventive measures before the actual emergence of equipment problems, greatly improving the rapidity and accuracy of equipment maintenance.

V. Conclusions

Industrial big data acquisition and cleaning technology is an important foundation for industrial digital transformation and intelligent manufacturing, which can efficiently collect, clean, and analyze the massive data in the industrial production process, thus providing powerful support for enterprise decision-making, production optimization, and quality improvement. This paper researches the key technology of big data collection and cleaning in cigarette factories, and proposes a big data collection and cleaning solution applicable to cigarette factories. The effectiveness of the program is verified through practical application cases. In the future, with the continuous development of technologies such as artificial intelligence, large models, and edge computing, the big data collection and cleaning technology of cigarette factories will be more mature and perfect, providing more powerful power for the digital transformation of the tobacco industry.

References

[1] Yang S.B., He H., Zhai Y.C., et al. (2024) Research on intelligent factory construction and its key technology based on big data. Electrotechnology 10:67-70. 10.19768/j.cnki.dgjs.2024.10.019.

[2] Hao W. (2024) Research on big data platform applied to internet of things. Communication World, 31(10):187-189. 10.3969/j.issn.1006-4222.2024.10.063.

[3] Wang T.T. (2024) Industrial big data collection and application in industrial internet environment. Modern Industrial Economy and Informatization,14(10):52-53. 10.16525/j.cnki.14-1362/n.2024.10.016.

[4] Wang P,Peng X.T. (2023) Exploring the application of industrial big data in industrial digital transformation. Intelligent Manufacturing,04:70-73.10.3969/j.issn.1671-8186.2023.04.015.

[5] Sun Q.B., Liu J., Li G.F., et al. (2010) Internet of things:A review of research on concepts, architectures and key technologies. Journal of Beijing University of Posts and Telecommunications 33(3):1-9. 10.3969/j.issn.1007-5321.2010.03.001.

[6] Fan J.H. (2023) Design and realization of IoT device management platform based on microservice architecture. University of Electronic Science and Technology.10.27005/d.cnki.gdzku.2023.001893.

[7] Zhang W. (2023) Exploration of industrial big data acquisition design for cigarette engineering trial production platform based on time series data. Changjiang Information and Communication,36(4):21-23. 10.3969/j.issn.1673-1131.2023.04.007.

[8] Liu F, Pan K. (2021) Generalized architecture analysis of industrial internet big data application. Electronic Testing,21:94-96.10.3969/j.issn.1000-8519.2021.21.033.

[9] Liang Z.Y.,Wang H.Z. (2023) A review of research on key technologies for time-series data analysis in smart IoT. Intelligent Computers and Applications,13(12):1-8.10.3969/j.issn.2095-2163.2023.12.002.

[10] Wan Z.Y. (2023) Design and realization of data processing platform and task scheduling system for edge IoT. Zhejiang University. 10.27461/d.cnki.gzjdx.2023.000136.

[11] Cao Y., Yu H. (2023) An intelligent cleaning algorithm for IoT temporal big data based on importance calculation. Automation and Instrumentation,12:71-75+80. 10.14016/j.cnki.1001-9227.2023.12.071.