Research on Text Data Pool of Intelligent Manufacturing for Plate Parts

Guangrong Li¹, Yuanguo Cao², Sufang Zhao², Yongqiang Bao³ and ChaoYu¹

1 School of Mechanical Engineering, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China
2 School of Automotive and Rail transit, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China
3 School of Information and Communication engineering, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China
Email: grongli@njit.edu.cn, caoyuanguo@hotmail.com, zhaosufang@njit.edu.cn, 785518297@qq.com, 846835565@qq.com

Abstract. Based on understanding of the structure and function of plate parts, these are analyzed including the data produced in the process of intelligent manufacturing and the characteristics and contents of the way of information expression in this paper. The data of the whole life cycle of plate parts is the research object combined with research status. The characteristics and functions of the text data pool are analysed. The framework of the text data pool for plate parts is constructed. The text data pool for intelligent manufacturing of plate parts is designed. A Hadoop platform is built by using Java, hive and other components. And the modules are built to achieve text data pool data fusion. A large number of data generated in the process of plate parts design and intelligent manufacturing can be analysed, processed and extracted more clearly, such as data collection, data format conversion, text disambiguation and other fusion functions, so as to achieve the purpose of data processing. These researches not only play an important role in data storage and processing, but also improve the efficiency and accuracy of plate parts design and intelligent manufacturing.

1. Introduction

In the intelligent manufacturing environment, big data has 4V characteristics of large quantity, diversity, low value density and rapidity [1]. These data come from network, smartphone, sensor, camera, payment system and other ways, which are important factors to improve productivity, innovation and competitiveness. Therefore, the research of big data is of great significance to intelligent manufacturing industry. As one of the most typical common parts, plate parts produce a lot of important heterogeneous data in the process of intelligent manufacturing. Different new technologies are applied to intelligent manufacturing, such as expert system technology, which can be used in the process design, production planning, fault diagnosis, etc. in the process of intelligent manufacturing. Artificial intelligence technology is more suitable for solving particularly complex and uncertain problems. This requires that intelligent manufacturing enterprises need more data calculation and processing. As a result, big data lake is a storage space containing multiple data pools, which can store a large number of data with different sources and formats. It classifies, integrates, extracts and utilizes the data generated in the production process, which will have a profound impact on the efficient production and future development of intelligent manufacturing enterprises. The text data pool is one of them, and data sources are the same and are integrated in the pool to form unified data.
format [2]. Therefore, the research of text data pool in intelligent manufacturing has important practical significance.

2. Research Status of Intelligent Manufacturing Data Pool for Plate Parts

Plate parts are one of the most common typical parts. In the process of producing plate parts, a lot of data will be generated from the initial design process to the whole stage of production and sales. Enterprises can also use the design parameters and processing parameters of plate parts to track the whole life cycle from design to manufacturing to working conditions in real time, so as to realize the intelligent production and design and the intelligent internal data sharing. Because the data pool covers all the data in the life cycle of the plate parts, it closely links the manufacturing workshop, department and the cooperative object of the enterprise, and avoids the isolated operation of each production and operation process of enterprise. The combination of intelligent manufacturing, information technology and processing design technology can realize the intelligent production of production beat, production speed and processing manufacturing of plate parts.

Figure 1. Plate part

In the content made in China 2025, intelligent manufacturing is positioned as the main direction of China's manufacturing industry to realize the transformation from big to strong and the deep integration of industrialization and informatization. In essence, it is consistent with the positioning of German industry 4.0, American industrial Internet, Japan's global intelligent manufacturing cooperation plan, and EU's IMS 2020 plan [3]. She Chengqi [4], et al. studied the application of big data in the new energy vehicle industry. This paper introduces the process of big data from data collection to data processing and data mining. Extensive heterogeneity and data quality are two key problems that big data technology needs to solve. M. Jebakumari [5] uses Hadoop's data storage and processing framework to manage the supply and demand of big data. Data Lake Management includes model (assist analysis), data asset catalog (ontology, taxonomy), workflow (process, schedule, capture), access management (based on group / role / authorization based on user), metadata. Durga Rao Karri et al. [6] summarized the relevant research of big data, introduced the composition of big data security framework, the concept of big data life cycle, summarized the opportunities and challenges of big data, analyzed the ideal structure of big data, and thought about the development direction of big data. Bozena Malysiak mrozek et al. [7] used fuzzy technology and u-sql query language (a big data processing language) to process the data stored in the big data lake, and designed and verified a series of analysis methods based on this technology. Chunhui Yang (2016) [8] designed a data fusion system, explaining the system structure from three aspects of data acquisition, calculation and storage. B. Rajendran (2016) [9] proposed to use the largest tree graph to express data and the KDD cup 99 multivariate data set of 40,000 instances with 42 features.
3. Intelligent Manufacturing Text Data Pool for Plate Parts

3.1. Contents and Characteristics of Intelligent Manufacturing Data for Plate Parts

In the intelligent manufacturing enterprise for plate parts, a plate part will involve a lot of text data information from design to sales, such as design dimension, processing technology, assembly dimension, test data and sales data. These text data information storage forms are different, and each kind of data has a very large amount of data, which makes intelligent manufacturing enterprise of the plate parts not to effectively process and use these text data. It is very important of how to store and analyse these text data after format standardization[10].

In the design process of plate parts, the design dimensions include nominal diameter DN, pipe outer diameter a, pipe diameter D, bolt hole center circle diameter K, bolt hole diameter L, thread TN, etc. These data are all in digital form, which is easy to be unified in data format. During the processing of plate parts, there are processing dimensions and production processing logs. At this time, the obtained text data not only has the form of numbers, but also has the form of text. It is difficult to unify the format of text data and to process data.

In addition, there will be other forms of text data types in the whole process of intelligent manufacturing of plate parts, such as the dimensions related to the matching properties between parts, the dimensions to ensure the relative positions between parts and the dimensions to be processed during assembly, etc., as well as the dimensional errors, end perpendicularity errors and coaxiality errors involved in the detection of plate parts.

3.2. Text Data Pool for Intelligent Manufacturing of Plate parts

Through analysis and summary, the data source of text data pool mainly has the following two ways, as shown in Figure 2.

1. The data of intelligent manufacturing for plate parts first enters the original data pool, and if it is text type data, it will be imported into the text data pool.
2. The data for intelligent manufacturing of plate parts is put into the simulation data pool and application data pool according to the type. After standardized processing, if the probability of future use is small, it is put into the text data pool.

3.3. Design of Text Data Pool for Plate Parts

Due to the different data sources of the text data pool for intelligent manufacturing of plate parts, the file formats entering the text data pool are different. There are semi-structured data such as doc format, xls format, SCV format, etc., structured data standardized database, and binary data such as audio, video and image, etc. These different forms of data are stored in the text data pool Medium[11] [12].

Data of different formats cannot be effectively processed in the text data pool, so it is necessary to convert files of different formats from the original data pool or the simulation data pool and the application data pool into a unified format in the text data pool. In the intelligent manufacturing text data pool for plate parts, the unified text file format of TXT is used to store, which can facilitate the text data pool to process and analyze these text data.
The big data lake for plate parts of intelligent manufacturing is divided into five different data pools, each of which has its own unique role. But the five data pools are closely related to each other. These data pools are logically differentiated, but do not require complete separation. When the data is standardized and analyzed in the intelligent manufacturing text data pool for plate parts, the text data will have different directions according to the future use. If the text data is often used after standardized processing and analysis, then the text data will be left in the intelligent manufacturing text data pool for plate parts. If the probability of using the text data in the future is very small, then the text data will be put into the intelligent manufacturing archive data pool for plate parts.

4. Data Processing in the Text Data Pool of Enterprise Intelligent Manufacturing for Plate Parts

The intelligent manufacturing enterprise of plate parts will produce a lot of data in the whole life cycle of plate parts, including design dimension data, production site data, detection data, storage data, sales data, etc. These data have the characteristics of large capacity and high complexity. Intelligent manufacturing can not do without the support of workshop production data. In the manufacturing process, CNC machine tool is not only the production tools and equipment, but also the node of workshop information network. Through the automatic collection, statistics, analysis and feedback of machine tool data, the results are used to improve the manufacturing process, which will greatly improve the flexibility of manufacturing process and the integration of processing process, so as to improve the quality and efficiency of production process[13] [14]. The field data such as unqualified product information and data collection of other related products is a big problem faced by intelligent manufacturing enterprises. The method which achieving efficient, simple and fast data collection is a problem that needs to be solved by intelligent manufacturing enterprises of plate parts. Now take the production log as the data source and analysis processing object to complete the following text data pool data collection.

4.1. Intelligent Manufacturing Data collection for Plate Parts

The significance of big data technology for intelligent manufacturing of plate parts does not lie in the use of large-scale data information, but in the intelligent processing of these data to analyze and mine valuable information, but the premise is to collect a large number of valuable data. For the plate parts intelligent manufacturing enterprise, the data information generated in production site is the first-hand information in the whole production cycle, because only by analyzing and processing the data generated in the production line every day, the problems can be found in time in the production process, issued and corrected in time. At the same time, only when you can understand the more accurate production capacity of production line, the scheduling plan of personnel and equipment can be ready to arrange, and a clear plan is made in the next production process, so that the order receiving will not exceed the capacity load during sales. There are several ways to collect data for intelligent manufacturing of plate parts: 1) real-time collection of output data from the production line or the number of defective products, or the type of failure of the production line (such as line stop, material shortage, quality), and transmission to the text data pool; 2) receiving information from other data pools: such as the data of processing production including plan information, material information, etc, 3) transferring the name and quantity information of defective products in the inspection station, 4) connecting the inspection instrument, and realizing the digitalization of inspection instrument. The data collection instrument automatically obtains the measurement data from measurement instrument, and records, analyzes and calculates, forms the corresponding various graphs, and automatically judges the measurement results, such as the beating measurement of machined parts, the tension meter pull and drawing of force curve, etc. If you want to process these data, the first thing you need to do is to upload the information file from the local to Hadoop platform [15] [16], and then use the data integration to do the corresponding data processing and analysis.

Local files are uploaded by write programs to Hadoop platform, and a "new" folder is created in Hadoop platform. The files are uploaded to the folder, and the file is renamed.
4.2. Upload the Text data Pool file for Intelligent Manufacturing of Plate Parts

In order to process the text data in the text data pool of plate parts intelligent manufacturing, it is necessary to upload the file to Hadoop platform. Hadoop implements a distributed file system (HDFS). It is a software framework that can deal with a large number of data in a distributed way. Hadoop works in a parallel way. It can speed up the processing speed through parallel processing with a very large capacity, and can reach PB level data.

The local host address is first set to "HDFS: //localhost: 9000" for upload files to HDFS of Hadoop, and then the storage folder "/ new" is created under the host Hadoop address. The code is "private static final string dir_path ="/ new ";. Then a new file " / newfile" is created under the folder, code "private static final string file" path = " / new / newfile"; ". Finally, the local file is read and written to the new file. The code for uploading the local file is "InputStream in = new FileInputStream ("C:\hadoop-2.7.7\\YC / d.txt");".

The main process of uploading local files to Hadoop platform is as following: creating a new folder in Hadoop platform → creating a TXT file → reading and writing the created TXT file to the local file.

After the text data in the text data pool for intelligent manufacturing of plate parts is processed, the files generated from the processing results can be automatically uploaded to the Hadoop platform. If the processing results are needed to view, the results file can be downloaded from Hadoop platform. When the intelligent manufacturing enterprise for plate parts uses Hadoop platform, only the enterprise is a unified network, then it can log in the same HDFS page for file upload and download, which can facilitate the operation of personnel who need data processing. Hadoop has the following advantages: 1) high reliability, 2) high scalability, 3) high efficiency, 4) high fault tolerance, 5) low cost. These advantages are enough to make intelligent manufacturing enterprises choose to use Hadoop platform, which can optimize their production development, improve manufacturing process, improve production efficiency and improve product quality.

4.3. Cleaning of Redundant Data for Intelligent Manufacturing of Plate Parts

For the data generated in the process of intelligent manufacturing of plate parts, the quantity is very large, because the production log will be updated every day. It is necessary to clean up the backup data after being processed and analyzed every day in time. Next, a program is written for cleaning the timeout data to check the date of the local backup file. If it is more than 24 hours, the file is deleted.

First, the local file is made by a file format date definition as "yyyy MM DD HH" type. If the file creation date is 6:00 on May 12, 1999, the file date type is "1999-05-12-06". Then the algorithm of "long now = new date (). Gettime()" is used to get the date of the local text data file. Then an IF statement is used to judge the backup time of these acquired date files. If the backup time is more than 24 hours, the backup file will be deleted in the local backup folder. The if statement code is "if (now time > 24 * 60 * 60 * 1000L) {FileUtils. Deletedirectory (DIR)}". In this way, after the data cleaning program runs, it will traverse the folders which are cleaned up, and then the files are deleted if the backup file times out.

5. Realization of the Text Data Pool of Enterprise Intelligent Manufacturing for Plate Parts

A Hadoop platform is designed and built, and some components are used, such as Java, MapReduce, hive. According to the way of data conversion, a text data pool is built, and the data of plate parts is processed. Some interfaces are shown in Figure 3. According to the data information of plate parts, on the basis of building the text data pool, data processing is carried out. The text data pool stores the data from text, image, audio and other data sources, and stores it on HDFS of Hadoop in the format of TXT. It also uses MapReduce to process the original data, which is convenient for the data processor to process, and realizes the functions of data text disambiguation, data conversion and so on.
6. Summary
This paper analyzes the data information in the whole life cycle of intelligent manufacturing process of plate parts. The text data pool for plate parts is established based on Hadoop and other big data technologies, realizing the processing of data information from the text data source, application data pool and simulation data pool. It provides accurate and detailed data support for the intelligent manufacturing of plate parts, and also realizes the high-efficiency production for the enterprise. It can save time and cost, so as to achieve the highest benefit. Further research work will be carried out in the next step, such as semantic judgment, semantic conflict and conflict resolution of text data.

7. Acknowledgments
This project is supported by the youth project of humanities and social science research fund of the ministry of education, China (No. 17YJCZH083), higher education reform project of Jiangsu province, China (No. 2019JSJG139), the project of humanities and social science research fund of the ministry of education, China (No. 16YJCZH108).

8. References
[1] Liu Zhihui, Zhang Quanling. Overview of big data research technology, Journal of Zhejiang University, 2014.
[2] LIANG Zhiyu,WANG Hongzhi,LI Jianzhong,GAO Hong.A Review on the Application of Big Data Analysis in Manufacturing Industry, Mechanic,2018, vol.6, PP.1-13.
[3] XIAO Jinghua, MAO Yunchi, XIE Kang. Transformation of Chinese Manufacturing Industry: Insight from Intelligent Manufacturing System Based on Internet and Big Data. Industrial Economic Review, 2016,(02),P7-18.
[4] SHE Chengqi, ZHANG Zhaosheng, LIU Peng, SUN Fengchun. Overview of the Application of Big Data Analysis Technology in New Energy Vehicle Industry: Based on Operating Big Data of New Energy Vehicle. JOURNAL OF MECHANICAL ENGINEERING,2019,55(20),p3-12.
[5] M.Jebakumari. Current Trends in Building and Managing Data Lakes [J]. Volume 7, Issue 3 (May 2018)
[6] Durga Rao Karri, Kr Koteswara Rao. The Data Management with Big data and its Major Concepts [J]. International Journal of Mechanical Engineering and Computer Applications, 2015, 3(6), p379-384.
[7] Bozena Malysiak-Mrozek, Marek Stabla, Dariusz Mrozek. Soft and Declarative Fishing of Information in Big Data Lake [J]. IEEE TRANSACTIONS ON FUZZY SYSTEMS, 2018, 26(5), p2732-2747.

[8] Chunhui Yang, Bo Cheng. Architecture of Data Fusion System Based on Big Data Technology. International Conference on Mechatronics Engineering and Information Technology (ICMTEIT 2016).

[9] B. Rajendran1 and Saravana Venkataraman2. ENSEMBLE FUZZY SUPPORT VECTOR MACHINE CLASSIFIER BASED ON MAXIMUM SPANNING TREE FOR BIG DATA ANALYTICS. ARPN Journal of Engineering and Applied Sciences. 2016, 11(9).

[10] Z Yan, J Liu, L.T. Yang, N Chawla. Big data fusion in Internet of Things [J]. Information Fusion 40 (2018)

[11] Antonis Troumpoukis1, Jurgen Jakobitch2, and Vangelis Karkaletsis. Stasinos Konstantopoulos1, Angelos Charalambidis1, Giannis Mouchakis1, Semantic Web Technologies and Big Data Infrastructures: SPARQL Federated Querying of Heterogeneous Big Data Stores. http://cassandra.apache.org/doc/cql3/CQL.html 2015, p1-4.

[12] Ajit Singh1, Sultan Ahmad. Architecture of Data Lake. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN: 2456-3307. DOI https://doi.org/10.32628/CSEIT1952121

[13] T. Roja1, A. Murali Mohan Kumar. A Review on Data Fusion and Integration. International Journal of Computer Sciences and Engineering. Vol.7, Special Issue, 6, March 2019, pp: 77-81.

[14] M. Jebakumari. Current Trends in Building and Managing Data Lakes. International Journal of Scientific Research in Computer Science Applications and Management Studies. Volume 7, Issue 3 (May 2018) ISSN 2319 – 1953:

[15] R. Tan, R. Chirkova, V. Gadepally, and T. G. Mattson. Enabling Query Processing across Heterogeneous Data Models: A Survey. In 2017 IEEE International Conference on Big Data (BIGDATA 2017), Boston, USA, p 3211–3220, December 2017.

[16] Anne Shepherd, Chalermporn Kesa, James Cooper, Joel Onema, Paul Kovacs. OPPORTUNITIES AND CHALLENGES ASSOCIATED WITH IMPLEMENTING DATA LAKES FOR ENTERPRISE DECISION-MAKING. Issues in Information Systems, Volume 19, Issue 1, pp. 48-57, 2018.