Big data analytics for intelligent manufacturing systems: A review

Article *in* Journal of Manufacturing Systems · June 2021

DOI: 10.1016/j.jmsy.2021.03.005

CITATIONS
442

READS
9,478

4 authors:

Junliang Wang
Donghua University
94 PUBLICATIONS **2,165** CITATIONS
SEE PROFILE

Xu Chuqiao
Shanghai Jiao Tong University
23 PUBLICATIONS **938** CITATIONS
SEE PROFILE

Jie Zhang
Donghua University
155 PUBLICATIONS **3,443** CITATIONS
SEE PROFILE

Ray Y. Zhong
The University of Hong Kong
312 PUBLICATIONS **17,588** CITATIONS
SEE PROFILE

Journal of Manufacturing Systems xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Big data analytics for intelligent manufacturing systems: A review

Junliang Wang [a,b,c], Chuqiao Xu [d], Jie Zhang [a,b,*], Ray Zhong [e]

[a] Institute of Artificial Intelligence, Donghua University, Shanghai, China
[b] Shanghai Engineering Research Center of Industrial Big Data and Intelligent System, Donghua University, Shanghai, China
[c] National Synthetic Fiber Engineering Technology Research Center, Beijing Chonglee Machinery Engineering Co., Ltd, Beijing, China
[d] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China
[e] Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Big data analytics (BDA)
Intelligent manufacturing
Artificial intelligence
Manufacturing systems

ABSTRACT

With the development of Internet of Things (IoT), 5 G, and cloud computing technologies, the amount of data from manufacturing systems has been increasing rapidly. With massive industrial data, achievements beyond expectations have been made in the product design, manufacturing, and maintain process. Big data analytics (BDA) have been a core technology to empower intelligent manufacturing systems. In order to fully report BDA for intelligent manufacturing systems, this paper provides a comprehensive review of associated topics such as the concept of big data, model driven and data driven methodologies. The framework, development, key technologies, and applications of BDA for intelligent manufacturing systems are discussed. The challenges and opportunities for future research are highlighted. Through this work, it is hoped to spark new ideas in the effort to realize the BDA for intelligent manufacturing systems.

1. Introduction

With the arrival of a new round of industrial revolution, information technology has accelerated its integration with manufacturing systems, and the data owned by enterprises have become increasingly rich, with features of volume, variety, and velocity [1]. In intelligent manufacturing, industrial big data not only promotes enterprises to accurately perceive the internal and external environment changes in the system but also facilitates scientific analysis and decision making to optimize the production process, reduce costs and improve operational efficiency. With massive data, new business modes are generated out of expectation (such as mass customization [2] and precision marketing [3]) to empower the social development and economic growth. As a result, big industrial data is regarded as a mean of production to driving intelligent manufacturing.

With the development of artificial intelligence, big data analytics (BDA) have been greatly improved to effectively mine both structured and unstructured industrial data in intelligent manufacturing, which has become a new research hotspot [4]. Continuous learning from big data of manufacturing system enables the system to be self-learning, self--optimization and self-regulation [5]. With the further development of BDA, the operation of manufacturing systems will be deeply changed [6–8]. To understand the current state of the research and provide insights for future studies, this paper systematically analyzes current research efforts, derive prominent research themes, and identify gaps in the current literature about big data analytics for intelligent manufacturing systems (BDAIMS).

A bibliometric analysis was conducted with published data from 2011 to 2020 regarding intelligent manufacturing driven by big data gathered from the Web of Science database (as shown in Fig. 1), which shows an almost steady increase in papers on this topic. Fig. 1(a) shows the number of published articles about intelligent manufacturing from 2011 to 2020. From 2011–2014, the amounts of articles are small and have little change. But from 2014 to 2019, the amounts have been increased sharply and the ratio keeps going up. Affected by COVID-19, the amounts of relevant articles published in 2020 have decreased. Fig. 1(b) shows the top areas related to intelligent manufacturing. As shown in the chart, the top five areas are the Engineering, the Computer Science, the Automation Control Systems, the Business Economics and the Telecommunications. The Computer Science is the most related area for intelligent manufacturing. With the technology developing rapidly, the Computer Science will be the key to break through the bottleneck of intelligent manufacturing. Fig. 1(c) shows the productive scholars publishing in this area. Fig. 1(d) lists the top universities or research

* Corresponding author at: Institute of Artificial Intelligence, Donghua University, Shanghai, China.
E-mail address: mezhangjie@dhu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.jmsy.2021.03.005
Received 30 November 2020; Received in revised form 4 March 2021; Accepted 5 March 2021

institutes publishing in this area. The top five universities are South China University of Technology, Chinese Academy of Sciences, Shanghai Jiao Tong University, University of Hong Kong and Huazhong University of Science Technology. This figure shows that most of them are Chinese universities, which corresponds to the following analytics about countries or regions. As shown in Fig. 1(e), countries or regions that are active in this field, of which China, the United States, and the United Kingdom are the top three. These articles are sourced from the Web of Science database with a focus on key concepts of intelligent manufacturing and big data. By analyzing these key technologies and related academic movements, the way forward is clearer and the next new wave is coming soon.

2. Concepts and definitions

2.1. Concept of big data

With the development of IoT, intelligent manufacturing has focused on the collection of enormous data called big data [6]. However, it is facing great challenges when making full use of such data. The continuous manufacturing process, various sensing devices and real-time and efficient data transmission make the data possess the typical characteristics of "3V" of big data of volume, variety and velocity [7]. Through further analysis of the data in manufacturing workshops, it is found to be characterized by multi-source, multi-dimension, multi-noise, imbalance, and time series.

Multi-source: Along with the rise of IoT technologies and the number of sensors in workshops, the manufacturing data is stored and captured in multi-sourced information systems [8] with a different data structure [9]. To integrate multi-source data in the manufacturing process, Zhu [10] proposed a big data analytics framework for smart tool condition monitoring (TCM), using multi-source information fusion method to process image data, 3-d point cloud data, and frequency signal data, achieve the monitor and adaptively control of machining process in real time under varying working conditions. Zhang [11] proposed a pixel level fusion method that combines raw data from multiple sources into single resolution data, and summarizes several trends tending to broaden the application of multi-source data fusion. Tao [12] analyzed the dependency of data in product lifecycle management under big data environment, and proposed a conceptual

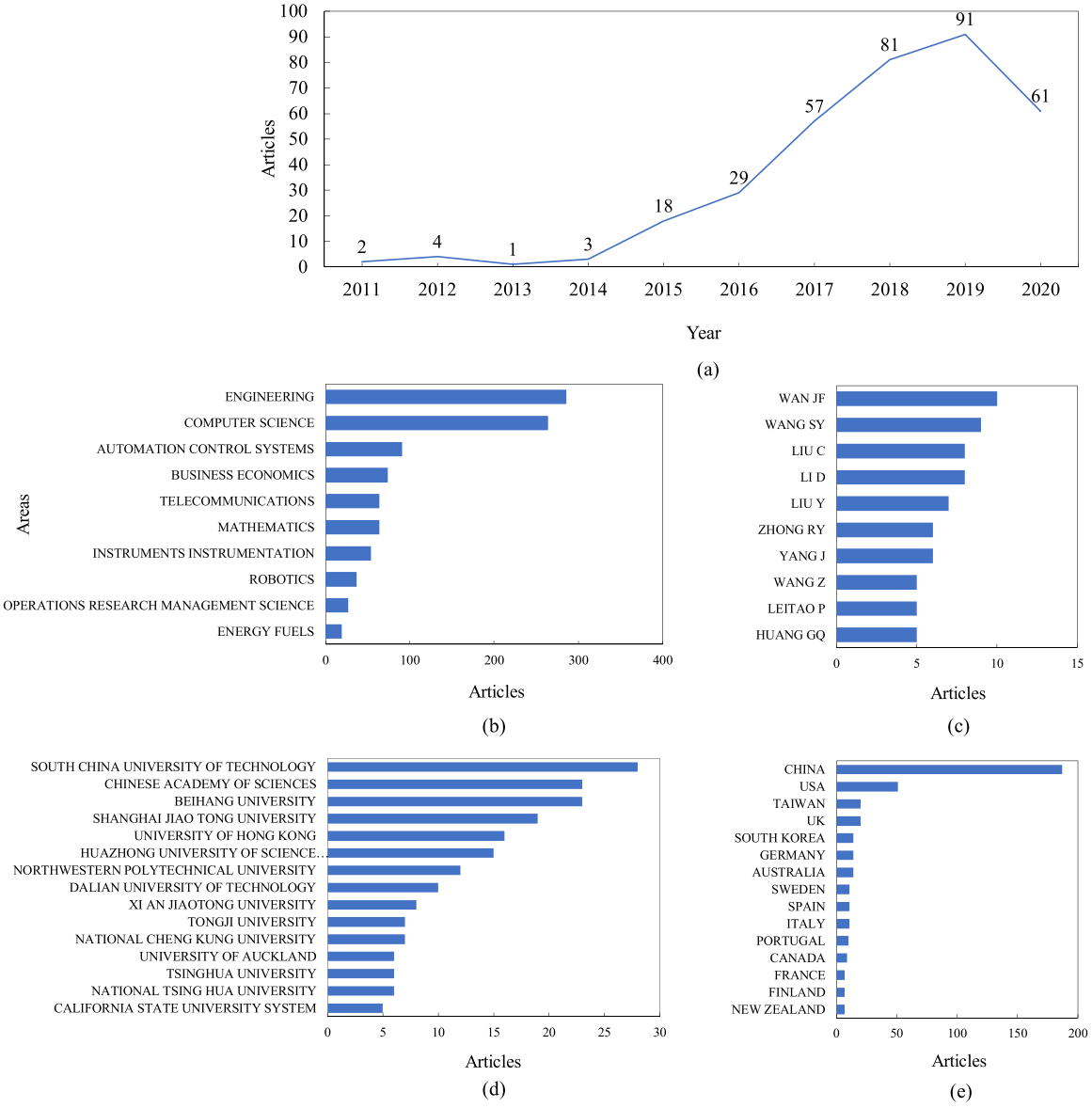


Fig. 1. Statistics from Web of Science database (search keywords: “intelligent manufacturing” and “big data”; Date: 25 November 2020). (a) Published articles per year; (b) published articles by areas; (c) published articles by author; (d) published articles by affiliation; (e) published articles by country/region.

framework with higher flexibility, accuracy, and less computing time, which can deal with multi-source data and massive data. Zhang [13] proposed an overall architecture of multi-source lifecycle big data for product lifecycle (BDA-PL), to make better product lifecycle management (PLM) and (cleaner production) CP decisions in manufacturing processes.

Multi-dimension: In manufacturing systems, the performance parameters are influenced by different factors, which results in a multi-dimension problem in prediction and control [14]. The current methods for multi-dimension problems can be classified into numerous types. For example, in semiconductor wafer manufacturing systems, the cycle time of wafer products is influenced by more than one thousand factors, such as the processing time of each operation, the size of waiting queue for each machine, and the utilization of a machine [15]. In health condition monitoring of the machines, to fully inspect the health conditions, condition monitoring systems are used to collect real-time data from multiple sensors after the long-time operation [16]. Aiming at the multi-dimension feature of manufacturing data, Liu [17] proposed a novel integrated process planning and control method based on intelligent software agents and multi-dimension manufacturing features. Luo [18] proposed a cloud manufacturing (CMfg) architecture to process multi-dimension data, achieving on-demand use, dynamic collaboration, and circulation of manufacturing resources.

Multi-noise: Industrial big data also has the characteristics of multi-noise at the same time since the electromagnetic interference and harsh environment [19]. For example, in a dataset of wafer cycle time forecasting, about five thousand records are containing missing value, and fifteen hundred records containing abnormal value, in every 8 million records [20]. To solve the tasks with multi-noise data, Lee [21] proposed a systematic and data-driven approach to process these missing values, and random sampling data for identifying semiconductor wafers defects. In order to reduce the impact of noise data, Zhao [22] developed the deep residual networks with dynamically weighted wavelet coefficients (DRN + DWWC) to adaptively filter out these multi-noise data, which improved the accuracy of gearbox fault diagnosis.

Imbalance: The number of samples in the industry also has the characteristic of imbalance. For example, the data samples with defects are much less than normal ones [23]. As indicated by the previous BDA studies, the class imbalance is a critical issue accounting for the bad performance, since they are biased toward the majority classes and tend to misclassify minority class examples [24]. In pattern recognition problems, several approaches have been proposed to solve the classification tasks with imbalanced data, which can be classified into two types: model modifying method, and data processing method [25]. Kang and Zhou [26] proposed a distance based weighted down-sampling scheme for support vector machine (SVM) for imbalanced classification. On the opposite, the oversampling method synthetic data instances to augment the minority classes to change the distribution of classes. Zhang et al. [27] designed a weighted minority oversampling (WMO) to rebalance the data distribution during fault diagnosis of rotating machinery.

Time series : A lot of industrial data are collected in time series, such as the vibration signal, and temperature data. In wafer manufacturing production, the machine processing data, wafer measurements, defect distribution data, wafer electrical properties data, actual production monitoring data and machine operation status data are all time series data. These time series data are very important to provide variation tendency for improving the accuracy of prediction [28]. To deal with the time series data, Vafeiadis [29] proposed a self-adaptive sliding data window method to extract time series features. To process the captured data from sensors (time-series data) for predictive maintenance of equipment, Villalobos [30] presented a system I4TSRS1, analyzing time-series data by combining the time-series reduction techniques with extracted features from industrial time-series. In order to predict the quality of products in textile production, Seçkin [31] proposed an improved random forest algorithm for time series data. In this algorithm, the random decision trees were created via modified algorithm of bagging and unbiased split selection based on conditional probabil probability index so as to construct random forests. In petroleum production, there are a large amount of nonlinearity and complexity of time-series data, Sagheer [32] introduced a deep learning approach capable to address the limitations of traditional forecasting approaches and improve the accuracy of predictions.

2.2. Concept of model driven and data driven

There are two main paradigms for BDA: model-driven, and data-driven modes.

Model-Driven is the way to start with a solid idea of how the physical system works. Model-driven approaches are powerful because they rely on a deep understanding of the system or process, and can benefit from scientifically established relationships [33]. Models can't accommodate infinite complexity and generally must be simplified. They have trouble in constructing a well fitted model accounting for noisy data and non-included variables. In real practice scenarios, building up a sophisticated model integrates the physical, mechanical, electronic, data flow, or other appropriate details of the complex systems. Furthermore, modeling takes time. It is inherently a trial-and-error approach, rooted in the old scientific method of theory-based hypothesis formation and experiment-based testing. Finding a suitable model and refining it until it produces the desired results is often a lengthy process [34].

Data driven is another model-free way based on the correlation between system status parameters and the target estimated by carious artificial intelligent models, which have high accuracy for accurate optimization [35]. Find an algorithm that can spot connections and correlations by extracting knowledge from a database that can be obtained from historical measurements. Validations reveal their excellent performance in computing efficiency, accuracy, and decision-making rationality. However, data driven methods strongly rely on expert experience in designing machine learning algorithms and their performance is also affected by the quantity and quality of the prior-knowledge database. Data driven methods work based on the co-distribution hypothesis. The machine learning tools that discover features and train-up classifiers learn from examples, and there need to be enough examples to cover the full range of expected variation and null cases. To train the generalized model and discover viable feature sets and decision criteria, big data is truly needed to get meaningful results. The data driven methods training with big data are called to be big data driven.

3. Framework and development

3.1. The framework of big data analytics for intelligent manufacturing systems

The primary science paradigm of BDAIMS is data science, which was highlighted in the book entitled the fourth paradigm: data intensive scientific discovery. In 2009, Tony et al. pointed out from the perspective of scientific research paradigm that data intensive science would become the next paradigm after experimental science, theoretical derivation and simulation [36]. The traditional scientific research paradigms construct sophisticated mathematical models to approximate real systems through experiment, derivation and simulation, and analyzes and optimizes the system. In complicated large-scaled dynamic systems, well experimental, mathematical, and simulation models are hard to be constructed. On the other hand, big data extracts knowledge by mining the correlation between data, which can provide stronger insight, analysis and decision-making ability.

Under the paradigm of data science, the big data driven operation of manufacturing systems change to emerge a framework of "correlation + prediction + regulation" (shown in Fig. 2). (1) correlation analysis

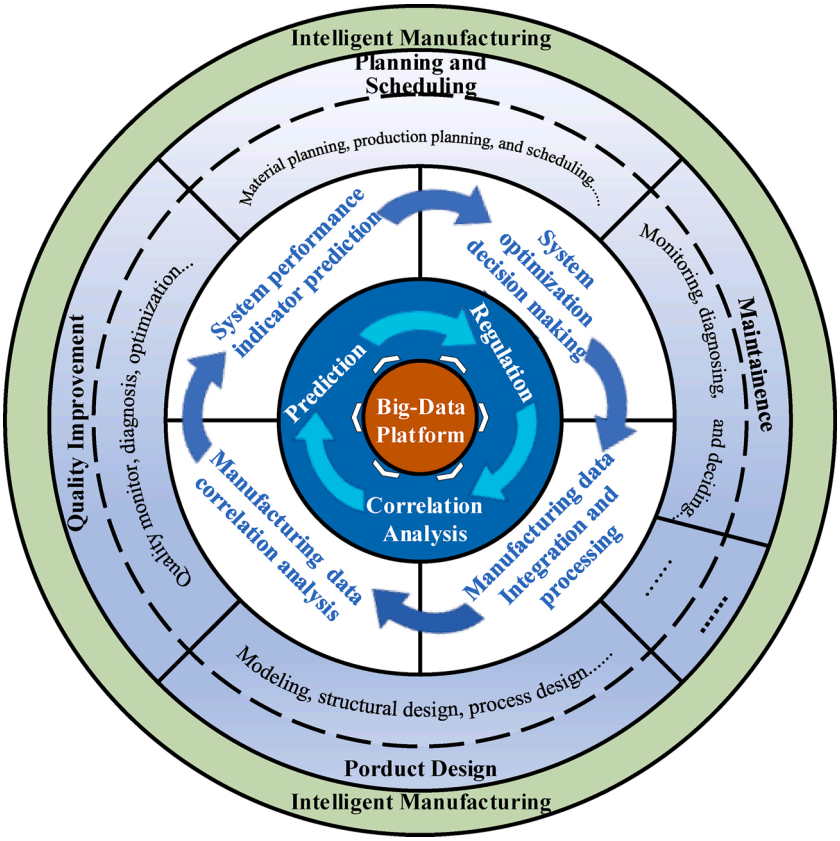


Fig. 2. The framework of big data driven intelligent manufacturing (adjusted from [37]).

means to quantify the relationship between different factors in the manufacturing systems from the perspective of data. (2) Prediction refers to further forecast the performance indicators of manufacturing systems (e.g. cycle time and yield) by machine learning methods. (3) Regulation refers to optimizing the controllable variables to improve the system performance. From the view of data processing cycle, the BDAIM works in four steps: (1) The manufacturing data is integrated and preprocessed to provide reliable and reusable data. (2) The correlative analysis is conducted to obtain the explanatory factors of performance indicators of manufacturing systems. With the explanatory factors, the fluctuation of system performance indicators can be modeled. (3) With the explanatory factors, the system performance indicators can be predicted to provide insights for decision making. Different kinds of machine learning models are developed for accurate prediction. In this section, prediction is generalized to consist fault detection, classification, and other extended prediction tasks in manufacturing systems. (4) Based on the predicted value, the decision-making methods can be implemented to improve the systems performance. Usually, the function, structure, process of product can be optimized by design data analysis. The efficiency of manufacturing systems can be optimized through planning and scheduling. The yield of products was controlled and improved by process and quality control systems. The system robustness was guaranteed by prognostics health management in intelligent maintain.

3.2. The development of big data analytics

The big data analytics are enabled by the evolution of artificial intelligence [38], which can be divided into three generations as shown in Fig. 3.

1) The first generation of BDA

The first generation of BDA aims to solve complex problems with simple algorithms. The representative work was influenza epidemics detector from google [39]. This detector estimated the probability that a random physician visit in a particular region was related to an ILI indicator which was equivalent to the percentage of ILI-related physician visits. In this detector, a simple linear regression model was developed to model the relationship between an ILI physician visit and the log-odds of an ILI-related search query. With large numbers of Google search queries, the influenza-like illness in a population was accurately tracked for each of the nine surveillance regions of the United States without any influenza transmission models. In the first generation of BDA, data is treated as the most important ingredient to be a success.

2) The second generation of BDA

The second generation of BDA aims to solve a complex problem with complicated algorithms. Deep learning models are central enablers for this generation. The massive parameters (such as weights and biases) in the deep neural networks can be adjusted to handle high dimensional problems. In this period, outstanding progress has been made in the complicated high dimensional problems, such as prognostics health management of equipment, surface defect recognition, neural language processing. The representative mark was the AlphaGo designed by Deep Mind [40], which introduced Monte Carlo search tree and deep learning models into Go games. The latest version AlphaGo Zero was trained by self-play during 4.9 million games for approximately 3 days. However, the number of games in the life of a human player could be ten thousand class. With the power of training and deep learning algorithms, no human professional player can beat AlphaGo Zero. In manufacturing systems, different kinds of problems can be modeled and solved by various kinds of deep learning methods. For example, the recurrent neural network has been integrated with convolutional neural networks to tackle the fault detection problems with time series data in the

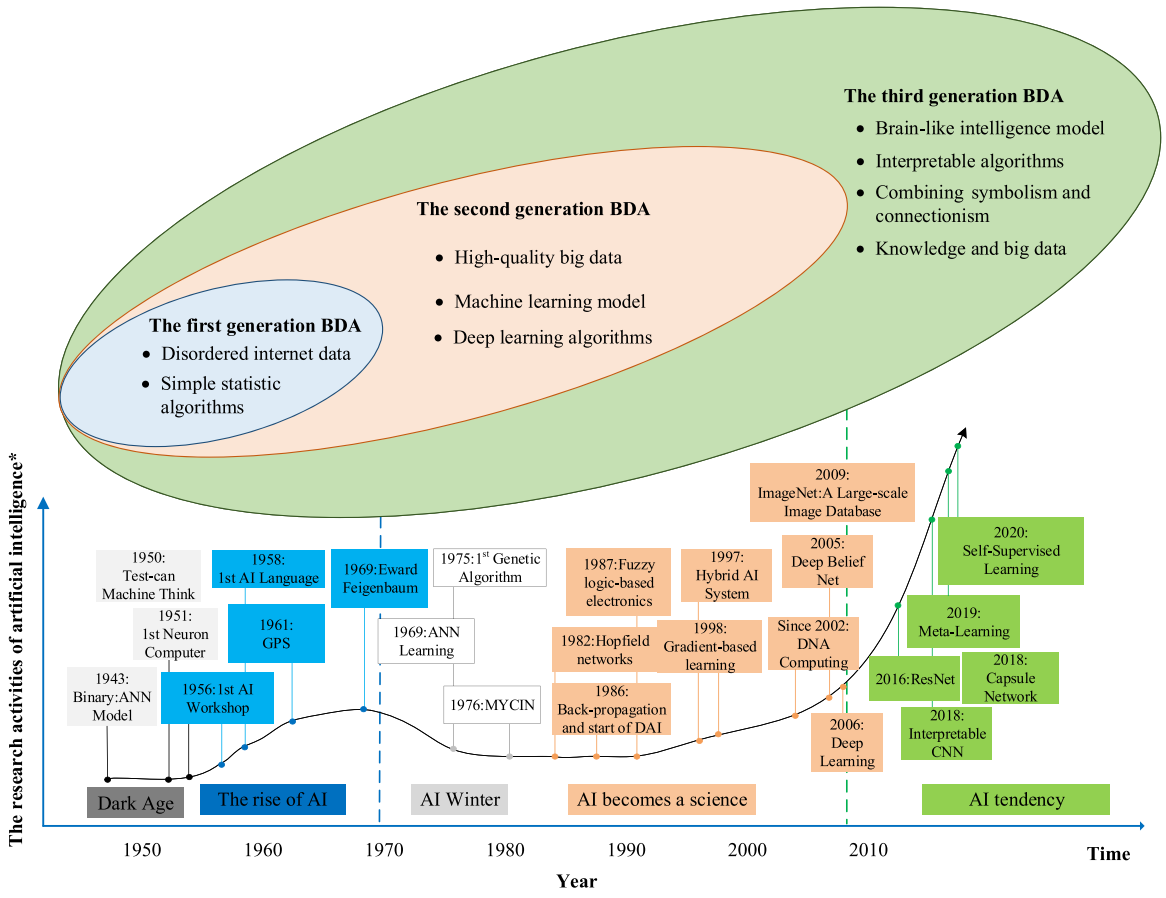


Fig. 3. The development of big data analytics (adjusted from [38]).

intelligent maintain of a machine.

In the second generation of BDA, well designed complex models and high-quality big data are the fundamental issues to succeed in manufacturing systems. Nevertheless, the need for generalization and interpretability represent substantial obstacles to the development and translation of BIDA. The first important challenge in developing realizable and efficient systems is access to large, specific and well-annotated datasets. Deep learning is unique in its capability of recognizing meaningful patterns from raw data without explicit directions when provided with sufficient examples. Current published BDA works on the co-distribution hypothesis, where the distribution of training and testing data should be identical. However, the data collected from industrial practice is changing and unpredictable since manufacturing systems are dynamic evolutionary. The second important challenge is that the inner workings and decision-making processes of BDA algorithms remain opaque. For the security considerations, the industrial applications require decision support tools to explain the rationale or support for its decisions to enable the users to independently review the basis of their recommendations. To meet this requirement and gain trust from industrial users, BDA should provide explanations for their outputs. Overcoming these two major barriers to industrial implementation can facilitate the development and adoption of BDA into industrial practice.

3) The next generation of BDA

The next generation of BDA is placed great expectations to approach the level of human intelligence with the development of AI. Zhang, one of the pioneers of AI in China, regards the integration of knowledge, data, algorithm, and computation as the core driving force of the next generation of BDA [41]. He thought the combination of symbolism and connectionism is imperative to establish a new, explainable, robust AI theory and develop safe, trust worthy, reliable, and extensible AI technology. Hassabis et al. [42] argued that better understanding of biological brains could play a vital role in building intelligent machines. To gain insights from bio-science for AI has become a consensus view. Shimon Ullman [43], director of Weizmann AI center, considered bio-structure as a critical success factor. He thought combining deep learning with brain-like innate structures may guide network models toward human-like learning. In industrial practice, the BDA model is learnt and evaluated during the interaction with the dynamic environments. The mechanisms and development of model structure topology and behavior may be an important issue to empower the next generation of BDA [44].

4. Key technologies of big data analytics

4.1. Computation framework

With the development of data sensing and Internet of Things technologies, the scale of data continues to increase to reach TB, PB, and even higher levels, which can no longer be completed by a single computer. This puts forward higher requirements for the computing power and timeliness of the computer to process big data. Big data processing involves the configuration of the storage system, the division of computing tasks, the distribution of computing load, the data migration between computers, and the data security when the computer or network fails. The situation is much more complicated. Distributed computing technology, in simple terms, is that multiple machines share computing tasks, which is the main solution at present. Furthermore, in some industrial scenarios, due to the high real-time data calculation and analysis requirements of a large number of processes and equipment at the bottom of the production process, the edge-cloud computing

architecture was born.

1) Distributed technology

Around 2004, three papers introducing distributed file system GFS (Google file system) [45], parallel computing model (MapReduce) [46], and non-relational data storage system (BigTable) [47] were published by Google. This was the first time to propose a reusable solution for the distributed processing of big data. Inspired by the ideas of Google, Yahoo engineers Doug and Mike developed Hadoop. Based on improving Hadoop, dozens of big data computing frameworks applied in distributed environments have been born, as shown in Fig. 4.

(1) The first generation of distributed computing framework

The first generation of distributed computing framework represented by Hadoop initially mainly included two parts: Hadoop distributed file system (HDFS) and the computing framework (MapReduce) [48]. In version 2.0, the resource management and task scheduling functions are stripped from MapReduce to form YARN [49], so that other frameworks can also run on Hadoop like MapReduce. Compared with the previous distributed computing framework, Hadoop hides many tedious details, such as fault tolerance, load balancing, etc., making it easier to use [50]. MapReduce can be understood as combining a bunch of chaotic data according to certain characteristics, and then processing and obtaining the final result, as shown in Fig. 5. The basic processing steps are as follows:

a) The input file is divided into pieces according to certain standards, and each piece corresponds to a map task. In general, MapReduce and HDFS run on the same set of computers, that is, each computer is responsible for storage and calculation tasks at the same time, sharding usually does not involve data replication between computers.
b) The content in the fragment is usually parsed into key-value pairs according to a predefined certain standard.
c) Perform map tasks by processing each key-value pair, and then outputting key-value pairs.
d) MapReduce obtains the grouping method defined by the application and sorts the key-value pairs output by the map task according to the grouping. Each key name forms a group by default.
e) After all nodes have performed the above steps, MapReduce starts the Reduce task. Each group corresponds to a Reduce task.
f) The process executing the reduce task obtains all the key-value pairs of the specified group through the network.
g) Combine values with the same key name into a list.
h) Perform reduce task by processing the list corresponding to each key, and then outputting the result.

(2) The second generation of distributed computing framework

It can be seen that through the combination of multiple MapReduce, complex calculation problems can be expressed [51]. However, the combination process requires manual design and the process is relatively cumbersome. In addition, all computers need to be synchronized at each stage, which affects execution efficiency [52]. In order to overcome the above problems, a directed acyclic graph (henceforth referred to as DAG) calculation model is proposed [53]. The core idea of DAG is to decompose the task internally into a number of sequential subtasks, which can express various complex dependencies more flexibly. Microsoft Dryad [54], Google FlumeJava [55], and Apache Tez [56] are the earliest DAG models. Dryad defines several simple DAG models such as series connection, full connection, and fusion, and describes complex tasks by combining these simple structures. FlumeJava and Tez form DAG model by combining several MapReduces.

Another shortcoming of MapReduce is the use of disks to store intermediate results, which seriously affects the performance of the system, which is more obvious in situations that require iterative calculations such as machine learning. Spark, developed by the AMP laboratory of the University of California, Berkeley, overcomes the above problems [57]. Spark has improved the early DAG model, and proposed a memory-based distributed storage abstract model resilient distributed dataset (henceforth referred to as RDD), which selectively loads and resides in the intermediate data in the memory, reducing the IO overhead of disks. Compared with Hadoop, memory-based Spark operations are more than 100 times faster, and disk-based operations are also more than 10 times faster.

(3) The third generation of distributed computing framework

Machine learning and artificial intelligence are also one of the trends in big data computing. Spark and Flink launched machine learning libraries Spark ML and Flink ML, respectively. More platforms provide machine learning on third-party big data computing frameworks, such

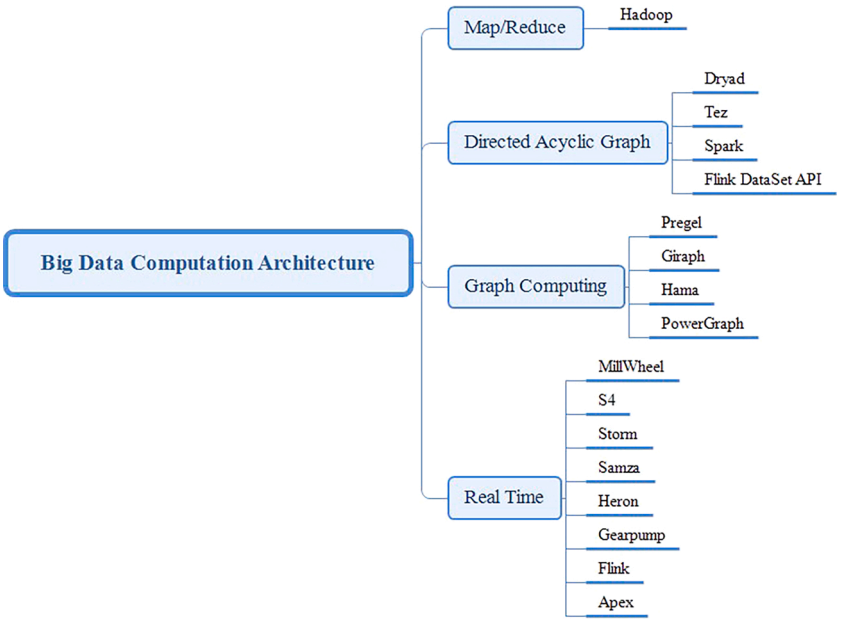


Fig. 4. Classification of distributed computing frameworks.

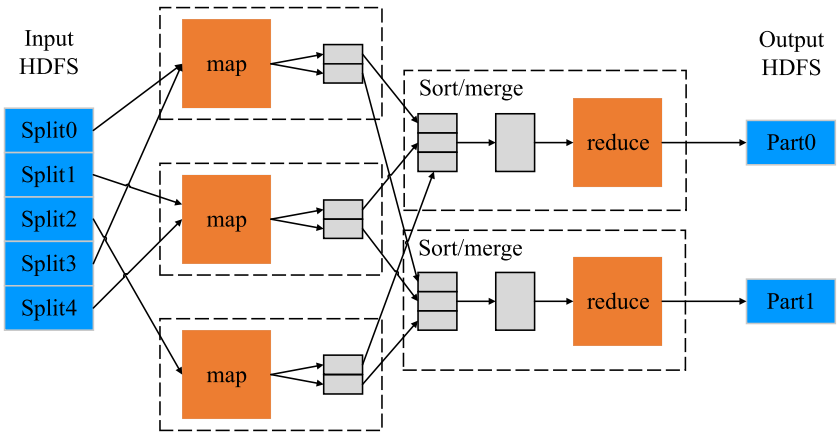


Fig. 5. MapReduce processing process (adjusted from [48]).

as Mahout, Oryx and the Apache incubation project SystemML, HiveMall, PredictionIO, SAMOA, MADLib. These machine learning platforms generally support multiple computing frameworks at the same time. For example, Mahout uses Spark, Flink, and H2O as engines, while SAMOA uses S4, Storm, and Samza. After deep learning has caused extensive discussion, some communities have explored the combination of deep learning frameworks with existing distributed computing frameworks, such as projects such as SparkNet, Caffe on Spark, and TensorFrames [58].

Supporting multiple frameworks on the same platform is also one of the development trends, especially for communities with strong development capabilities. Spark takes the batch processing model as its core, and implements the interactive analysis framework Spark SQL, the stream computing framework Spark Streaming (and the currently implemented Structured Streaming), the graph computing framework GraphX, and the machine learning library Spark ML. While Flink provides low-latency stream computing, batch processing, relational computing, graph computing, and machine learning keep running, and the goal is to go straight to the big data general computing platform. Google's Beam is trying to incorporate computing frameworks such as Spark, Flink, and Apex under its own standards, which with the intention of taking a monopoly on the market.

2) Cloud- Edge computing technology

With the increasing pace of the manufacturing process, the timeliness requirements for data capturing and analysis are getting higher and higher. Industrial applications such as real-time monitoring of the

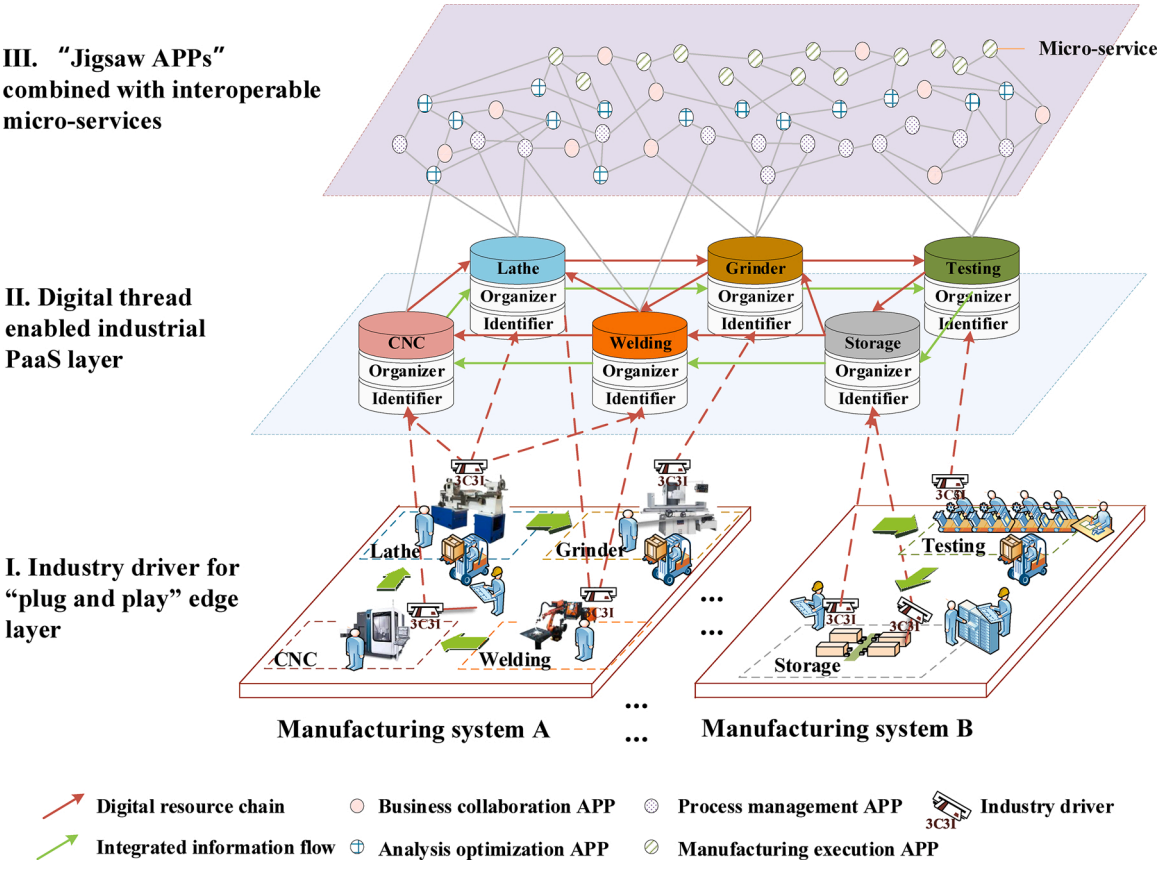


Fig. 6. The framework of "plug and play" edge computing model (adjusted from [59]).

production process, online quality inspection, and decision-making control all have high requirements for the timeliness of computing. Under the high timeliness requirements of industrial applications, the traditional way of data capturing by sensors, realizing data transmission through industrial Ethernet, and realizing data analysis instruction issuance in the enterprise private cloud cannot meet the requirements. Edge computing, which is optimized by edge nodes, has obvious advantages in low-latency applications. It is a powerful supplement to the cloud computing system.

(1) "plug and play" edge computing

Wang et al. [59] proposed a novel "plug and play" edge computing mode, as shown in Fig. 6, which consisted of the edge layer, the PaaS layer, and the application layer. In the special edge layer, each resource unit in workshop equipped with an industry driver for better accessibility. The industrial driver is an agent that contains the corresponding hardware and software acting as an abstraction layer between the edge and cloud. Benefited by the great computing power from the cloud, the industrial driver contained "3C-3I" structures: communication, computation, control, identification, insight and interoperation. Different from the driver in operation systems, the driver of a resource unit had functions of computing, and insight, which enabled the industrial drivers to monitor, analysis, and forecast the status of resource units to achieve initiative, efficient optimization with low time latency.

(b) Integration of 5 G network and edge computing

The 5 G network provides an environment in which edge computing can be widely deployed, and has the advantages of ultra-low communication delay and high positioning accuracy. Therefore, the combination of 5 G network and edge computing may become a typical scenario of edge-cloud convergence. Meanwhile, the unprecedented complexity of 5 G network also brings challenges to its integration with edge computing. This complexity comes from many aspects, such as dense network functions, heterogeneous, and highly diversified applications. Therefore, the deployment of edge computing in 5 G networks has become an important research area, and comprehensive solutions need to be developed to meet the challenges of heterogeneity, scalability, flexibility. Duan et al. [60] proposed a decentralized structure for service orchestration/federation across admin-domains in the 5 GEx and 5 G-Transformer projects, as shown in Fig. 7.

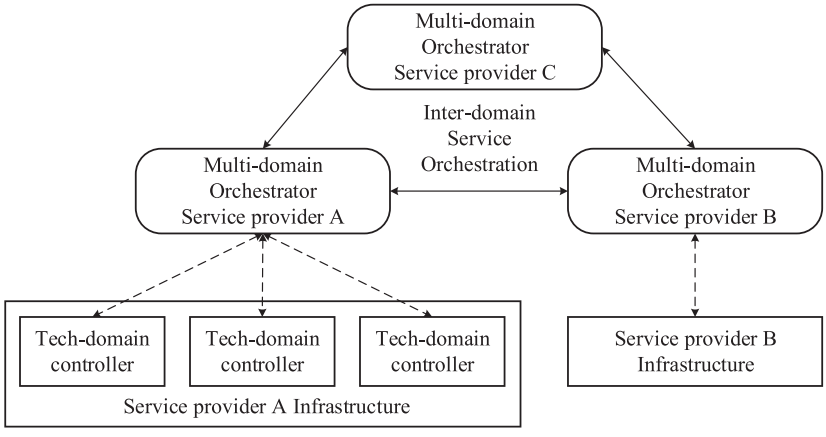
4.2. Data processing technology

The importance of big data is not how much data you have, but what you do with it. We can perform high-performance analysis on the massive production process data to find answers to reduce costs, shorten time, develop new products and optimize products, and make smart decisions. The mainstream data processing technology can be divided into batch processing and stream processing in terms of data processing mode. The characteristics of batch processing and stream processing were summarized and compared [61], as shown in Table 1.

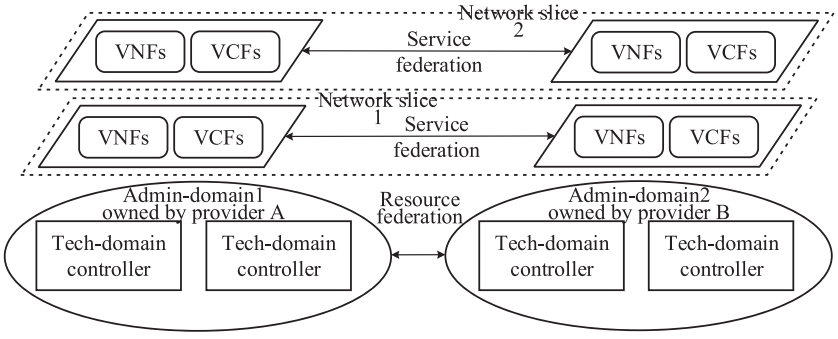
1) Batch processing framework

Table 1
Batch processing vs stream processing (adjusted from [61]).

Characteristics	Batch Processing	Stream Processing
Data scope	Queries or processing over all or most of the data in the dataset.	Queries or processing over data within a rolling time window, or on just the most recent data record.
Data size	Large batches of data.	Individual records or micro batches consisting of a few records.
Performance	Latencies in minutes to hours.	Requires latency in the order of seconds or milliseconds.
Analyses	Complex analytics.	Simple response functions, aggregates, and rolling metrics.



(a) Inter-domain service orchestration in 5G-Exchange



(b) service / resource federation across admin-domains in 5G-Transformer

Fig. 7. Decentralized structure for service orchestration/federation across admin-domains in the 5 GEx and 5 G-Transformer projects (adjusted from [60]).

Batch processing mainly operates on large static data sets and returns the results after the computation is completed. Batch task can be interpreted as a collection of data points that have been grouped at a fixed time interval (each piece of data is called a data point and corresponds to the state information of a point in time). We can also call this a window of data.

The data sets used in bath processing mode always conform to these characteristics:

a) Bounded: A batch data set represents a finite set of data.
b) Persistent: Data is usually always stored in some type of persistent storage location.
c) Vast: Batch operations are often the only way to process extremely large data sets.

Batch processing is well suited for computations that require access to a complete set of records. For example, when calculating totals and averages, the data set must be treated as a whole rather than as a collection of multiple records. These operations require the data to maintain its own state as the calculation proceeds.

Tasks that require processing large amounts of data are often best handled by batch operations. Whether the data set is processed directly from the persistent storage device or loaded into memory first, the batch processing system is designed with the amount of data in mind to provide sufficient processing resources. Because batch processing is particularly good at handling large amounts of persistent data, it is often used to analyze historical data.

The processing of large amounts of data requires a lot of time, so batch processing is not suitable for situations with high processing time requirements. Batch tasks are suitable for the following situations:

a) Aggregate function operations are needed to run on data over a period of time, such as the average, maximum, and minimum values of these data points.
b) The alert function does not need to run every data point, or the state represented by the data, and does not change frequently.
c) Reduce the frequency of collection to a large extent (downsample). Only the most obvious part of the huge data is needed to explain the change of the overall state.
d) A little extra latency won't have a big impact on your overall business.
e) Kapacitor processes data slower than it writes data when the cluster is running the timing database.

Apache Hadoop and its MapReduce processing engine provide a proven batch processing model that is best suited for handling very large data sets with low time requirements [62]. A fully functional Hadoop cluster can be built with very low-cost components, making this cheap and efficient processing technique flexible for many cases. Compatibility and integration with other frameworks and engines make Hadoop the underlying foundation for a variety of workload processing platforms using different technologies.

2) Stream processing

In the era of big data, data is usually generated continuously and dynamically. In many cases, data needs to be processed in a very short period of time, and fault tolerance, congestion control and other issues need to be considered to avoid data omission or double-counting. Stream processing is the solution to this kind of problem. DAG (directed acyclic graph) model is generally used in stream processing framework. The nodes in the figure are divided into two categories. One is the input node of data, which is responsible for interacting with the outside world and providing data to the system. The other is the computing node of the data, which is responsible for performing certain processing functions such as filtering, accumulating, merging, etc. Continuous real-time data from external systems streams through these nodes and concatenates them. If the data stream is compared to water, the input node is like the sprinkler head, which keeps pouring water out, and the computing node is like the interface of the water pipe.

In the stream computing framework, Storm [63] is one of the most popular and widely used. This is due to Storm's simple programming model and support for Java, Ruby, Python, and other development languages. Storm also has good performance, handling millions of messages per second on multi-node clusters. Storm's solution is to assign an ID to each message as a unique identifier and include the ID of the original input message in the message. The state of each original input message is maintained with a response center (Acker) with an initial value of the ID of the original input message. After the successful execution of each computing node, the ids of the input and output messages are xor, and then xor the state of the corresponding original input message. Since each message is xor once respectively during generation and processing, all messages are xor twice after successful execution, and the corresponding state of the original input message is 0. Therefore, when the state is 0, the content of the original input message can be safely cleared. If the state is not 0 after the specified time interval, it is considered that there is something wrong with the processing of the message and needs to be re-executed.

3) "Batch + stream" processing

With the further development of big data, the separate batch processing and stream processing framework could not fully meet the current needs of the enterprise, hence the hybrid mode combined batch processing and stream processing was born, as shown in Fig. 8.

Apache Spark [64] is the typical framework for "batch + stream" processing. Spark is an optimization based on the Hadoop MapReduce computing model. Spark greatly improves the processing capacity of data through the in-memory computing model and execution optimization (in different cases, the speed can be 10–100 times faster than MR or even higher). However, Spark's Streaming capability is provided by Spark Streaming module. Spark introduced the concept of Micro-batch, which treats access data over a short period of time as a single micro-batch. However, compared with native stream processing systems such as Storm, Spark Streaming has a relatively high latency.

Apache Flink [65] also supports stream processing and batch processing. Flink's design idea is "stateful stream processing", which treats input data item by item as a real stream and batch tasks as a bounded stream. In the current field of streaming data processing framework, Flink is unique. While Spark also provides batch and stream processing capabilities, the micro-batch architecture of Spark stream processing makes its response time slightly longer. Flink stream-first approach achieves low latency, high throughput and true item-by-item processing, which is why Flink has received more and more attention in recent years.

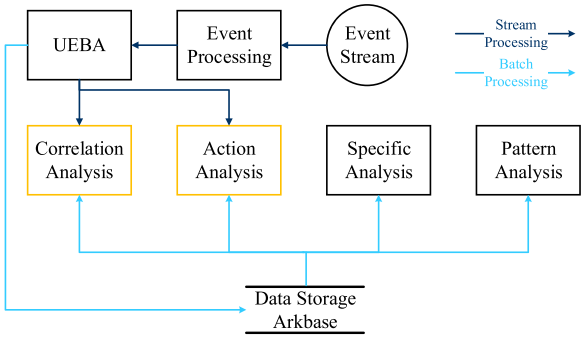


Fig. 8. "Batch + stream" processing.

5. Application in manufacturing systems

Driven by intelligent sensing, Internet of Things (IoT), distributed storage computing, machine learning and other technologies, big data-driven intelligent manufacturing applications have begun to emerge. As is shown in Fig. 9. the applications in product design, planning and scheduling, quality optimization, equipment operation and maintenance have been carried out to substantialize the intelligent manufacturing systems.

5.1. Big data-driven product design

Design plays an important role in manufacturing process, which determines most of a product's performance and performance [12]. With the development of big data technology, product design is shifting towards data-driven design from subjective conceptual design [66]. Big data-driven product design analyzes the market demand through users' evaluation [67]. Then, it generates the design scheme through the sale, customer, manufacturing data, and enhances the decision-making ability based on history learning. Finally achieves the closed-loop optimization of self-monitoring, self-analyzing, self-learning and self-adaptation [37]. Intelligent design can be divided into two aspects according to the design goals, one is product design for quality improvement, and the other is humanized shape design focused on customer preferences.

In the field of product design, a new digitized and parameterized perspective are shaped in the era of big data. The main factors affecting the quality can be obtained by analyzing various types of process data through correlation analysis, and established the mapping model of quality influencing factors to effectively optimize the product design parameters [68]. Tao [69] analyzed the isolation of data in the product life cycle management under the big data environment, and proposed a digital twin-driven model of product design, manufacturing and service. Geiger [35] analyzed the reliability of components in automobile design by using field measurement data and user usage data to improve design parameters and enhance product reliability. Tucker [70] proposed big data-driven product optimization method, in which the design decision tree model was used to realize the combinatorial optimization of products. Qin [71] applied big data analyzing technology to identify key parameters affecting diesel engine power by correlation analysis which effectively reduced the cost of design.

Big data provides a mechanism of fine design with streamlined design processes, promoted product innovations, and developed customized products by researching and understanding customer demands, behaviors, and preferences [72]. Ireland [73] brought Internet user evaluation into product design to achieve a quantitative analysis of product functions and provided support for design decisions. Xu [74] proposed a framework of data-driven product design for capturing product visual aesthetics UX to effectively identify the useful design concepts from consumer preferences to consumer response. Promotion strategies were developed to suit for corresponding segments of different customers. Li [75] designed a framework of smart product. In this framework, pairwise comparative algorithm was employed to cluster technical attributes into modules assembled into smart products. Wireless sensor network was developed to monitor smart product, and K-means algorithm was adopted to deal with monitoring data. Triangular fuzzy numbers were used to evaluate the maturity of the monitoring function. Finally, the design of control function was determined to give the optimal strategy by Q-learning algorithm to realize the closed loop optimization of design [76].

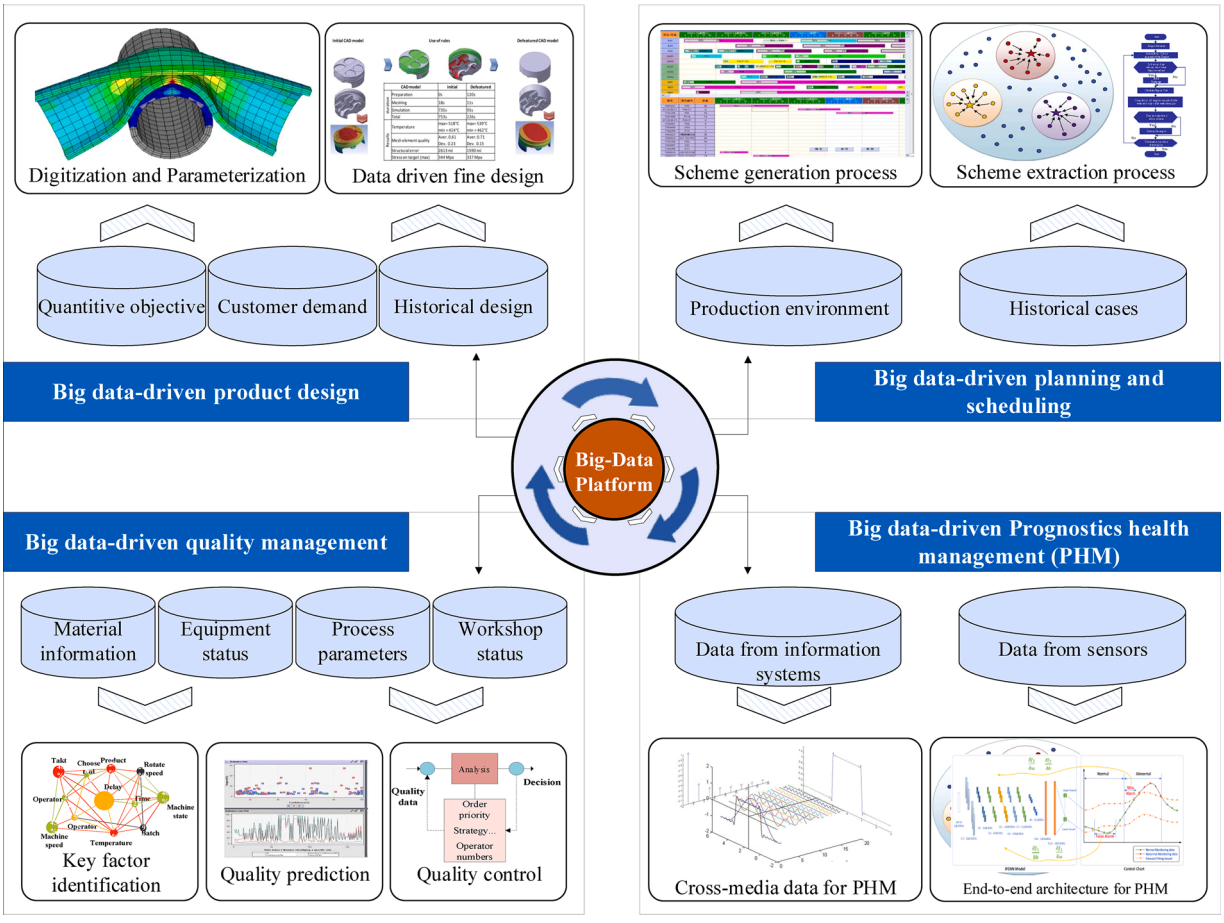


Fig. 9. The applications of BDA in manufacturing systems.

5.2. Big data-driven planning and scheduling

Production planning and scheduling is one of the core issues of enterprise operation and management [77]. In the ideal scenario, a feasible schedule can efficiently allocate production resources while coping with the impact of various on-site uncertainties. However, traditional shop floor scheduling methods mainly focus on the optimization of workshop performance and resource utilization in static environment, which may result in deterioration or infeasibility of static schedule. Big data analytics provide a series of effective methods and tools for the optimization of production planning and scheduling in dynamic environment [78].

BDA can improve the generation of scheme in the planning and scheduling. A large amount of historical data and real-time information of the production process should be collected through various enterprise information systems, embedded sensors and intelligent machines, such as the processing time of jobs, the setup time of machines, and the transportation time of materials [79–81]. Based on the collected data, big data analytics can provide comprehensive information support for scheduling decisions. For example, before making production schedules, the distribution information of various uncertain factors can be accurately estimated by statistical analysis techniques. After that, robust optimization [82,83] and intelligent optimization [84,85] algorithms are used to formulate robust production plans and schedules under uncertain environment.

In the execution process of a scheme, by extracting the current status information of workshop, such as manufacturing resource status, production progress and inventory data, deep learning and other machine learning methods [86] can be used to predict the completion time of products, so as to evaluate the production capacity. By establishing the risk probability evaluation model of schedules, high-risk tasks can be identified and reallocated [87]. Finally, the complete actual operation data and the execution results of dispatching rules provide useful knowledge about planning and scheduling, which can be used to establish reliable and practical heuristic scheduling rules based on learning strategies [88]. In addition, the production process can be optimized by analyzing the correlation of production parameters in historical data and the influence on production performance.

5.3. Big data-driven quality management

Big data-driven product quality management realizes product traceability and optimization based on product manufacturing process data and quality inspection data [89]. The main factors affecting product quality most, such as raw material performance parameters, equipment status parameters, process parameters, and workshop environmental parameters can be identified by correlation analysis. Meanwhile, a mapping model between quality influencing factors and quality performance can be established to effectively predict product quality. Intelligent optimization algorithms are further used to adaptively adjust control parameters that affect product quality in real time to achieve adaptive control and optimization of product quality.

Identifying the key influencing parameters affecting the product quality from mass manufacturing parameters plays an important role as the input of the product quality control model. Chien et al. screened out 12 highly correlated WAT factors through expert experience as model input, and designed an analysis method based on modified Partial Least Square (mPLS) to screen key parameters [90]. This method required the expert experience to select key parameters, which was difficult for quality analysts without experience to quickly master this skill. Qin et al. used bench test data in the quality control of diesel engines to perform correlation analysis on potential parameters that affect power consistency, and effectively improved power consistency [71]. These methods analyze the correlation between candidate factors and the quality indicators to reveal the root cause of fluctuation. However, the correlation evaluated by current BDA is different from the true situation under some conditions, since the observed correlation contains not only true dependence but also noises caused by transitive effects of correlations. As a result, the analyzed correlation may be heightened to mislead the key factor identification. To exactly identify the root cause factors, the causal inference embedded BDA could be further addressed to improve the robustness and accuracy of BDA.

In the self-adaptive optimization and control of product quality, BDA models of product quality are constructed to find the relationship between the raw collected data and product quality [91,92]. Gustavo et al. used data mining methods to predict 8 diesel quality performance parameters, which significantly reduced the detection time and cost [93]. Rokach L et al. applied data mining methods to improve the quality of the manufacturing process, and achieved good results in integrated circuit manufacturing [94]. Nakazawa et al. designed a five-layer convolutional network to achieve accurate recognition of 22 defect patterns on wafers [94]. Kiryong et al. designed a sub-fully connected layer network for each defect mode by sharing the convolution feature extraction layer. Each sub-network judged whether the defect mode existed in two categories, which effectively identified the defect mode of the wafer map [95]. Current methods performed well with complete, clean, and balanced data. The quality prediction or defect detection methods with imbalanced data [24], inequivalent data [96] is still a concern. In addition, the model is susceptible to input noise, and it is difficult to obtain high accuracy. For example, detecting product quality with noisy mixed- defects is a difficult issue [97]. At the same time, when the process parameters change, the special detection model that relies on data distribution training cannot dynamically adapt to the data change, resulting in model failure. Models in different case scenarios cannot be used universally so that the scene detection cannot be copied quickly, which influences the wide application of BDA in actual industries. The generalized BDA will be a next focused area of quality forecasting research.

Controlling product quality is an essential part of the development and production process. Jin et al. proposed a feedforward control method based on the picewise linear model [93]. The designed engineering-driven reconstruction method of the picewise linear regression tree further reduced the complexity of the model by merging leaf nodes under the constraints of control accuracy requirements. The effectiveness of the proposed method was verified in a multi-stage wafer manufacturing process. Borja et al. used Reinforcement Learning (RL) analysis solutions to derive the required feedback controller for the ball screw feed drive to provide positioning accuracy. The proposed method performed better than PID controller in computational experiments [98]. Liu et al. regarded the feature tensors extracted by CNN as interdependent feature sequences in the process of molten pool state recognition, and used LSTM to mine the relationship between the feature sequences to realize the effective integration of redundant features extracted by CNN [99]. Liu et al. designed a coarse-grained regularization method in view of the small sample problem faced in the process of molten pool state recognition based on deep learning, considering the characteristics of the convolution operation, which effectively prevented the model from overfitting [100]. Chien et al. used a neural network model to fit the relationship between WIP parameters, job size, the length of the waiting queue before the bottleneck station, the length of the waiting queue on the process route and the construction period to form a schedule control strategy [100]. Chen et al. used the Gauss-Newton regression model to measure the influence of the average number of layers, production capacity, movement, WIP and other parameters on the construction period, and realized the construction period control on this basis [101]. In the actual product manufacturing process, product quality often needs to be measured by multiple quality indicators, each of which is related to each other and involves multiple processes. The cooperative control with different variable controlled variables in the integration multi-process with design, manufacturing and assembly is still an open issue.

5.4. Big data-driven prognostics health management

Big data-driven prognostics health management (PHM) reveals the time-series changes in system fault characteristics through real-time monitoring of manufacturing process data, equipment performance parameters and other time-series data. The PHM data is fully analyzed to proactively identify potential anomalies in the system operation process in advance, diagnose the root cause of anomalies, and forecast the remaining life of a system [102].

The analyzed PHM data consists of acceleration signal, acoustic emission signals, optical signals, electrical data, temperature data, etc. In the PHM of a rotating machinery, Sinha et al. analyzed the vibration signals collected by an acceleration sensor [103]. For complicated machinery, Lei et al. acquired and integrated several acceleration sensors to monitor the health condition of planetary gearboxes [104]. In the PHM of high-speed moving machinery, Zhang et al. extended real-time nearfield acoustic holography (RT-NAH) to reconstruct the instantaneous surface normal velocity of a vibrating structure from the time-dependent pressure measured in the near field, which provides a non-contact and real-time method to measure and visualize the transient vibration of the structure [105]. Acoustic emission is another signal widely used in the PHM of slew bearing [106], cutting tool, and structure health monitoring applications. From the perspective of data, most of the current research using data within a single media. The cross-media BDA extracting information from different medias will be an important issue in the PHM.

Before the development of deep learning for BDA, the PHM methods make decisions based on pre-extracted features. There are several critical and commonly used features, including time-frequency domain-based feature, spectral cliff-based feature, higher-order statistical variable-based feature, and random resonance-based feature. The BDA merges two steps together, since deep learning methods can automatically extract features from raw time-domain signals [107]. Zheng et al. developed a transfer locality preserving projection based intelligent fault identification method, which embeds the data to a subspace through preserving a priori distribution structure properties and trains a classifier to identify the condition of target machine by the historical data [108]. To train the PHM model adaptively, Wen et al. developed a novel learning rate scheduler based on reinforcement learning for convolutional neural network in fault classification, which can schedule the learning rate efficiently and automatically [109]. Lei's group designed a feature-based transfer neural network to learn transferable features from data collected in a laboratory to achieve higher diagnosis accuracy for real-case machines [110]. Current BDA provides an end-to-end architecture to train PHM models with massive data directly. In industrial practice, the features, and potential abnormal signals, are changing with the dynamic operation environment. The BDA can discover knowledge about PHM with a self-learning mechanism to overcome the unpredictable scenarios. In the future, the end-to-end architecture may be improved with interpretable transformable quantizable learning technologies to substantialize a hybrid model.

6. Challenges for future research

6.1. Knowledge embedded industrial big data analytics

The BDA was considered as one of the most important technologies, due to its capacity to explore large and varied datasets to uncover hidden patterns and knowledge as well as other useful information [111]. The discovered patterns and knowledge can help manufacturing systems to make more-informed decisions, and to achieve the whole lifecycle optimization and more sustainable production [112]. The integration of current knowledge and the discovered patterns and knowledge from big data will be an interesting issue. From another perspective, the reuse discovered patterns and knowledge from big data into BDA process will improve the performance of BDA methods.

6.2. Cognitive security of big data analytics

The security of big data analytics in manufacturing systems is another major concern in the application. Although great achievements have been made, the overall performance of AI for BDA is sometimes surprising. The Google brain team in California observed an interesting phenomenon named antagonistic aggression, which describes that the current deep learning model is easy to be confused in an image classification task by added a small fluorescent sticker in the corner of the image [113]. Researchers at Kyushu University in Japan found that by changing only one pixel on an image, a neural network could be tricked during an image recognition task. Experiments on cifar-10 and ImageNet data sets achieved a deception success rate of 68.36 % and 41.22 %, respectively. This technique is therefore known as "One Pixel Attack". These cases essentially reflect that learning from massive data to improve the accuracy through black-box models is insufficient [114]. In manufacturing systems, the BDA is a part of the human cyber physical systems (HCPS), which requires high reliability and security. For BDA, it is essential to develop the cognitive security which ensures AI models operating properly from the attack of fake content, interference signals and other adversarial issues.

In this way the interpretability of machine learning models should be further concerned [115]. At present, deep neural networks obtain high discrimination power at the cost of low interpretability of their black-box representations [116]. In industrial scenarios, the transparency, and interpretability are required for security issues. The visualization of black-box models is a practical approach to understand the working process of deep learning methods. Another approach is based on statistical tools [117]. The sensitive analysis of different parameters and the role of the subcomponent can be investigated by knowledge distillation and network pruning [118].

6.3. Industrial big data governance

Governance of big data handles data integrity, quality, provenance, retention, processing, and analysis in full data lifecycle [119]. The governance of industrial big data considers the issues of security and privacy [120]. In the BDA, data from all sections are combined to form an integrated environment for system optimization [121]. In the majority of manufacturing organizations, raw data is not allowed to be transmitted to a remote commercial cloud center directly, since private raw data can easily be copied during storage and transmission. The Blockchain providing cybersecurity [122] may be a novel method for industrial big data governance. The architecture, mechanisms, technologies, and applications for big data governance keeping the privacy and security during big data sharing and integration is another open issue in the BDA.

7. Conclusion

As increasing attention is given to intelligent manufacturing systems, BDA is becoming a core technology to provide forecasting and decision making in manufacturing systems. BDA is a key future perspective in both research and industrial community, as it provides added value to various products and systems by applying cutting-edge technologies to traditional products in manufacturing. Key concepts, frameworks, key technologies, applications are discussed in this paper. Open issues for future research are highlighted after a systematic review. We hope this paper helps to promote the theoretical and technical research and engineering application of big data-driven intelligent manufacturing systems [123].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China under Grant 51905091, and in part by Shanghai Sailing Program under Grant 19YF1401500, Shanghai Science and Technology Planning Project under Grant 20DZ2251400. The authors want to thank Mr. Peng Zheng, Mr. Jie Ren, Mr. Shuxuan Zhao, Mr. Pengjie Gao, and Ms. Zixun Zhu for their help in literature collection.

References

[1] Gao RX, Wang L, Helu M, Teti R. Big data analytics for smart factories of the future. CIRP Ann 2020;69(2):668–92. https://doi.org/10.1016/j.cirp.2020.05.002.
[2] Leng J, et al. A loosely-coupled deep reinforcement learning approach for order acceptance decision of mass-individualized printed circuit board manufacturing in industry 4.0. J Clean Prod 2021;280. https://doi.org/10.1016/j.jclepro.2020.124405.
[3] Wang G, Gunasekaran A, Ngai EWT, Papadopoulos T. Big data analytics in logistics and supply chain management: certain investigations for research and applications. Int J Prod Econ 2016;176:98–110. https://doi.org/10.1016/j.ijpe.2016.03.014.
[4] Zhang Z, et al. Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. Nat Mach Intell 2019;1(May). https://doi.org/10.1038/s42256-019-0052-1.
[5] Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks : a review. Neural Netw 2019;113:54–71. https://doi.org/10.1016/j.neunet.2019.01.012.
[6] Kusiak A. Smart manufacturing must embrace big data. Nature 2017;544(7648). https://doi.org/10.1038/544023a.
[7] Zhong RY, Newman ST, Huang GQ, Lan S. Big Data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives. Comput Ind Eng 2016;101. https://doi.org/10.1016/j.cie.2016.07.013.
[8] Dutton W, et al. Clouds, big data, and smart assets: ten tech-enabled business trends to watch. McKinsey Q 2010;(4).
[9] Yager RR. A framework for multi-source data fusion. Inf Sci (Ny) 2004;163(1–3). https://doi.org/10.1016/j.ins.2003.03.018.
[10] Zhu K, Li G, Zhang Y. Big data oriented smart tool condition monitoring system. IEEE Trans Ind Inform 2019;16(6):1. https://doi.org/10.1109/tii.2019.2957107.
[11] Zhang J. Multi-source remote sensing data fusion: status and trends. Int J Image Data Fusion 2010;1(1). https://doi.org/10.1080/19479830903561035.
[12] Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. Int J Ind Manuf Syst Eng 2018;48:157–69. https://doi.org/10.1016/j.jmsy.2018.01.006.
[13] Zhang Y, Ren S, Liu Y, Si S. A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. J Clean Prod 2017;142. https://doi.org/10.1016/j.jclepro.2016.07.123.
[14] Wang J, Zheng P, Zhang J. Big data analytics for cycle time related feature selection in the semiconductor wafer fabrication system. Comput Ind Eng 2020;143(February):106362. https://doi.org/10.1016/j.cie.2020.106362.
[15] Wang J, Yang J, Zhang J, Wang X, Zhang W (Chris). Big data driven cycle time parallel prediction for production planning in wafer manufacturing. Enterp Inf Syst 2018;12(6):714–32. https://doi.org/10.1080/17517575.2018.1450998.
[16] Lei Y, Jia F, Lin J, Xing S, Ding SX. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans Ind Electron 2016;63(5). https://doi.org/10.1109/TIE.2016.2519325.
[17] Liu C, Li Y, Shen W. Integrated manufacturing process planning and control based on intelligent agents and multi-dimension features. Int J Adv Manuf Technol 2014;75(9–12). https://doi.org/10.1007/s00170-014-6246-0.
[18] Luo Y, Zhang L, Zhang K, Tao F. Research on the knowledge-based multi-dimensional information model of manufacturing capability in CMfg. Adv Mater Res 2012;472–475. https://doi.org/10.4028/www.scientific.net/amr.472-475.2592.
[19] Xu C, Wang J, Zhang J. Forecasting the power consumption of a rotor spinning machine by using an adaptive squeeze and excitation convolutional neural network with imbalanced data. J Clean Prod 2020:122864. https://doi.org/10.1016/j.jclepro.2020.122864.
[20] O'Donovan P, Leahy K, Bruton K, O'Sullivan DTJ. Big data in manufacturing: a systematic mapping study. J Big Data 2015;2(1). https://doi.org/10.1186/s40537-015-0028-x.
[21] Lee DH, Yang JK, Lee CH, Kim KJ. A data-driven approach to selection of critical process steps in the semiconductor manufacturing process considering missing and imbalanced data. J Manuf Syst 2019;52. https://doi.org/10.1016/j.jmsy.2019.07.001.
[22] Zhao M, Kang M, Tang B, Pecht M. Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes. IEEE Trans Ind Electron 2018;65(5). https://doi.org/10.1109/TIE.2017.2762639.
[23] He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng 2009;21(9). https://doi.org/10.1109/TKDE.2008.239.
[24] Wang J, Yang Z, Zhang J, Zhang Q, Chien W-TK. AdaBalGAN: an improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. IEEE Trans Semicond Manuf 2019;32(3). https://doi.org/10.1109/TSM.2019.2925361.
[25] Sun Y, Kamel MS, Wong AKC, Wang Y. Cost-sensitive boosting for classification of imbalanced data. Pattern Recognit 2007;40(12). https://doi.org/10.1016/j.patcog.2007.04.009.
[26] Kang Q, Shi L, Zhou MC, Wang XS, Di Wu Q, Wei Z. A distance-based weighted undersampling scheme for support vector machines and its application to imbalanced classification. IEEE Trans Neural Networks Learn. Syst 2018;29(9). https://doi.org/10.1109/TNNLS.2017.2755595.
[27] Zhang Y, Li X, Gao L, Wang L, Wen L. Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning. J Manuf Syst 2018;48(August 2017):34–50. https://doi.org/10.1016/j.jmsy.2018.04.005.
[28] Fernandez V. Wavelet- and SVM-based forecasts: an analysis of the U.S. Metal and materials manufacturing industry. Resour Policy 2007;32(1–2):80–9. https://doi.org/10.1016/j.resourpol.2007.06.002.
[29] Vafeiadis T, Krinidis S, Ziogou C, Ioannidis D, Voutetakis S, Tzovaras D. Robust malfunction diagnosis in process industry time series. IEEE International Conference on Industrial Informatics (INDIN) 2016. https://doi.org/10.1109/INDIN.2016.7819143.
[30] Villalobos K, Diez B, Illarramendi A, Go i A, Blanco JM. I4TSRs: a system to assist a data engineer in time-series dimensionality reduction in industry 4.0 scenarios. 2018. https://doi.org/10.1145/3269206.3269213.
[31] Seçkin M, Seçkin AÇ, Coşkun A. Production fault simulation and forecasting from time series data with machine learning in glove textile industry. J Eng Fiber Fabr 2019;14. https://doi.org/10.1177/1558925019883462.
[32] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 2019;323. https://doi.org/10.1016/j.neucom.2018.09.082.
[33] Benosman M. Model-based vs data-driven adaptive control: an overview. Int J Adapt Control Signal Process 2018;32(5):753–76. https://doi.org/10.1002/acs.2862.
[34] Wang Q, Li F, Tang Y, Xu Y. Integrating model-driven and data-driven methods for power system frequency stability assessment and control. IEEE Trans Power Syst 2019;34(6):4557–68. https://doi.org/10.1109/TPWRS.2019.2919522.
[35] Geiger C, Sarakakis G. Data driven design for reliability. Proceedings - Annual Reliability and Maintainability Symposium, 2016-April; 2016. https://doi.org/10.1109/RAMS.2016.7448023.
[36] Hey T, Tansley S, Tolle K. The fourth paradigm: data-intensive scientific discovery. Microsoft Research. 2009.
[37] Zhang J, Wang J, Lyu Y, Bao J. Big data driven intelligent manufacturing. Zhongguo Jixie Gongcheng/China Mech Eng 2019;30(2). https://doi.org/10.3969/j.issn.1004-132X.2019.02.001.
[38] Wang L. From intelligence science to intelligent manufacturing. Engineering 2019;5(4):615–8. https://doi.org/10.1016/j.eng.2019.04.011.
[39] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457(7232):1012–4. https://doi.org/10.1038/nature07634.
[40] Silver D, et al. Article mastering the game of Go without human knowledge. Nat Publ Gr 2017;550(7676):354–9. https://doi.org/10.1038/nature24270.
[41] Zhang B, Zhu J, Su H. Toward the third generation of artificial intelligence. Sci Sin Informationis 2020;50(9):1281–302. https://doi.org/10.1360/SSI-2020-0204.
[42] Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-inspired artificial intelligence. Neuron 2017;95(July (2)):245–58. https://doi.org/10.1016/j.neuron.2017.06.011.
[43] Ullman S. Using neuroscience to develop artificial intelligence. Science (80-) 2019;363(6428):692–3. https://doi.org/10.1126/science.aau6595.
[44] Rahwan I, et al. Machine behaviour. Nature 2019;568(7753):477–86. https://doi.org/10.1038/s41586-019-1138-y.
[45] Ghemawat S, Gobioff H, Leung ST. The google file system. Operating Syst Rev (ACM) 2003;37(5). https://doi.org/10.1145/1165389.945450.
[46] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Commun ACM 2008;51(1). https://doi.org/10.1145/1327452.1327492.
[47] Chang F, et al. Bigtable: A distributed storage system for structured data. ACM Trans Comput Syst 2008;26(2). https://doi.org/10.1145/1365815.1365816.
[48] Abouzeid A, Bajda-Pawlikowski K, Abadi D, Silberschatz A, Rasin A. HadoopDB: an architectural hybrid of mapreduce and DBMS technologies for analytical workloads. Proc VLDB Endow. 2009;2(1). https://doi.org/10.14778/1687627.1687731.
[49] Vavilapalli VK, et al. Apache hadoop YARN: yet another resource negotiator. 2013. https://doi.org/10.1145/2523616.2523633.
[50] Huang J, Nicol DM, Campbell RH. Denial-of-service threat to hadoop/YARN clusters with multi-tenancy. 2014. https://doi.org/10.1109/BigData.Congress.2014.17.
[51] Ranger C, Raghuraman R, Penmetsa A, Bradski G, Kozyrakis C. Evaluating MapReduce for multi-core and multiprocessor systems. 2007. https://doi.org/10.1109/HPCA.2007.346181.
[52] Stuart JA, Owens JD. Multi-GPU MapReduce on GPU clusters. 2011. https://doi.org/10.1109/IPDPS.2011.102.
[53] Bonifaci V, Marchetti-Spaccamela A, Stiller S, Wiese A. Feasibility analysis in the sporadic DAG task model. 2013. https://doi.org/10.1109/ECRTS.2013.32.
[54] Isard M, Budiu M, Yu Y, Birrell A, Fetterly D. Dryad: distributed data-parallel programs from sequential building blocks. Oper Syst Rev 2007:59–72. https://doi.org/10.1145/1272996.1273005.

[55] Chambers C, et al. FlumeJava: Easy, efficient data-parallel pipelines. ACM SIGPLAN Notices 2010;45(6). https://doi.org/10.1145/1809028.1806638.
[56] Saha B, Shah H, Seth S, Vijayaraghavan G, Murthy A, Curino C. Apache tez: a unifying framework for modeling and building data processing applications. Proceedings of the ACM SIGMOD International Conference on Management of Data, 2015-May; 2015. https://doi.org/10.1145/2723372.2742790.
[57] Zaharia M, Chowdhury M, Franklin MJ, Shenker S, Stoica I. Spark: cluster computing with working sets. 2010.
[58] Lyubimov D, Palumbo A. Apache mahout: beyond MapReduce. CreateSpace Independent Publishing Platform; 2016.
[59] Wang J, Xu C, Zhang J, Bao J, Zhong R. A collaborative architecture of the industrial internet platform for manufacturing systems. Robot Comput Integr Manuf 2020;61(April 2019):101854. https://doi.org/10.1016/j.rcim.2019.101854.
[60] Duan Q, Wang S, Ansari N. Convergence of networking and cloud/edge computing: status, challenges, and opportunities. IEEE Netw 2020;34(6). https://doi.org/10.1109/MNET.011.2000089.
[61] Shahrivari S. Beyond batch processing: towards real-time and streaming big data. Computers 2014;3(4):117–29.
[62] Inoubli W, Aridhi S, Mezni H, Maddouri M, Mephu Nguifo E. An experimental survey on big data frameworks. Future Gener Comput Syst 2018;86. https://doi.org/10.1016/j.future.2018.04.032.
[63] Christensen R, Wang L, Li F, Yi K, Tang J, Villa N. STORM: spatio-temporal online reasoning and management of large spatio-temporal data. Proceedings of the ACM SIGMOD International Conference on Management of Data, 2015-May; 2015. https://doi.org/10.1145/2723372.2735373.
[64] Zaharia M, Das T, Li H, Hunter T, Shenker S, Stoica I. Discretized streams: fault-tolerant streaming computation at scale. 2013. https://doi.org/10.1145/2517349.2522737.
[65] Carbone P, et al. Apache FlinkTM: Stream and Batch Processing in a Single Engine. 2015. p. 28–38.
[66] Kusiak A, Salustri FA. Computational intelligence in product design engineering: review and trends. IEEE Trans Syst Man Cybern Part C 2007;37(5). https://doi.org/10.1109/TSMCC.2007.900669.
[67] Qi J, Zhang Z, Jeon S, Zhou Y. Mining customer requirements from online reviews: a product improvement perspective. Inf Manag 2016;53(8). https://doi.org/10.1016/j.im.2016.06.002.
[68] Kusiak A. Innovation: a data-driven approach. Int J Prod Econ 2009;122(1). https://doi.org/10.1016/j.ijpe.2009.06.025.
[69] Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F. Digital twin-driven product design, manufacturing and service with big data. Int J Adv Manuf Technol. 2018;94(9–12). https://doi.org/10.1007/s00170-017-0233-1.
[70] Tucker CS, Kim HM. Data-driven decision tree classification for product portfolio design optimization. J Comput Inf Sci Eng 2009;9(4). https://doi.org/10.1115/1.3243634.
[71] Qin W, Zha D, Zhang J. An effective approach for causal variables analysis in diesel engine production by using mutual information and network deconvolution. J Intell Manuf 2020;31(7). https://doi.org/10.1007/s10845-018-1397-8.
[72] Wu D, Liu X, Hebert S, Gentzsch W, Terpenny J. Democratizing digital design and manufacturing using high performance cloud computing: performance evaluation and benchmarking. J Manuf Syst 2017;43. https://doi.org/10.1016/j.jmsy.2016.09.005.
[73] Ireland R, Liu A. Application of data analytics for product design: sentiment analysis of online product reviews. CIRP J Manuf Sci Technol 2018;23. https://doi.org/10.1016/j.cirpj.2018.06.003.
[74] Xu Z, Frankwick GL, Ramirez E. Effects of big data analytics and traditional marketing analytics on new product success: a knowledge fusion perspective. J Bus Res 2016;69(5). https://doi.org/10.1016/j.jbusres.2015.10.017.
[75] Li X, Ni Y, Ming XG, Song W, Cai W. Module-based similarity measurement for commercial aircraft tooling design. Int J Prod Res 2015;53(17). https://doi.org/10.1080/00207543.2015.1047973.
[76] Li X, Song W. A rough VIKOR-Based QFD for prioritizing design attributes of product-related service. Math Probl Eng 2016;2016. https://doi.org/10.1155/2016/9642018.
[77] Rossit DA, Tohmé F, Frutos M. Production planning and scheduling in Cyber-Physical Production Systems: a review. Int J Comput Integr Manuf 2019;32(4–5). https://doi.org/10.1080/0951192X.2019.1605199.
[78] Ismail R, Othman Z, Bakar AA. Data mining in production planning and scheduling: a review. 2009. https://doi.org/10.1109/DMO.2009.5341895.
[79] Zhong RY, Xu C, Chen C, Huang GQ. Big data analytics for physical Internet-based intelligent manufacturing shop floors. Int J Prod Res 2017;55(9). https://doi.org/10.1080/00207543.2015.1086037.
[80] Zhong RY, Huang GQ, Lan S, Dai QY, Zhang T, Xu C. A two-level advanced production planning and scheduling model for RFID-enabled ubiquitous manufacturing. Adv Eng Informatics 2015;29(4). https://doi.org/10.1016/j.aei.2015.01.002.
[81] Wang C, Jiang P. Manifold learning based rescheduling decision mechanism for recessive disturbances in RFID-driven job shops. J Intell Manuf 2018;29(7). https://doi.org/10.1007/s10845-016-1194-1.
[82] Zhang Y, Jin X, Feng Y, Rong G. Data-driven robust optimization under correlated uncertainty: a case study of production scheduling in ethylene plant. Comput Chem Eng 2018;109. https://doi.org/10.1016/j.compchemeng.2017.10.024.
[83] Zheng P, Zhang P, Wang J, Zhang J, Yang C, Jin Y. A data-driven robust optimization method for the assembly job-shop scheduling problem under uncertainty. Int J Comput Integr Manuf 2020. https://doi.org/10.1080/0951192X.2020.1803506.
[84] Wang B, Wang X, Xie H. Bad-scenario-set robust scheduling for a job shop to hedge against processing time uncertainty. Int J Prod Res 2019;57(10). https://doi.org/10.1080/00207543.2018.1555650.
[85] Gao D, Wang G-G, Pedrycz W. Solving fuzzy job-shop scheduling problem using DE algorithm improved by a selection mechanism. IEEE Trans Fuzzy Syst 2020. https://doi.org/10.1109/tfuzz.2020.3003506.
[86] Wang J, Zhang J. Big data analytics for forecasting cycle time in semiconductor wafer fabrication system. Int J Prod Res 2016;54(23):7231–44. https://doi.org/10.1080/00207543.2016.1174789.
[87] Ji W, Wang L. Big data analytics based fault prediction for shop floor scheduling. J Manuf Syst 2017;43:187–94. https://doi.org/10.1016/j.jmsy.2017.03.008.
[88] Jun S, Lee S, Chun H. Learning dispatching rules using random forest in flexible job shop scheduling problems. Int J Prod Res 2019;57(10). https://doi.org/10.1080/00207543.2019.1581954.
[89] Gao Y, GAO L, Li X. A generative adversarial network-based deep learning method for low-quality defect image reconstruction and recognition. IEEE Trans Ind Informatics 2020. https://doi.org/10.1109/tii.2020.3008703.
[90] Chien CF, Lee PC, Dou R, Chen YJ, Chen CC. Modeling collinear WATs for parametric yield enhancement in semiconductor manufacturing. IEEE International Conference on Automation Science and Engineering, 2017 August; 2017. https://doi.org/10.1109/COASE.2017.8256192.
[91] Yiping G, Xinyu L, Gao L. A deep lifelong learning method for digital-twin driven defect recognition with novel classes. J Comput Inf Sci Eng 2021;21(3). https://doi.org/10.1115/1.4049960.
[92] Miao R, Gao Y, Ge L, Jiang Z, Zhang J. Online defect recognition of narrow overlap weld based on two-stage recognition model combining continuous wavelet transform and convolutional neural network. Comput. Ind. 2019;112:103115. https://doi.org/10.1016/j.compind.2019.07.005.
[93] Jin R, Shi J. Reconfigured piecewise linear regression tree for multistage manufacturing process control. IIE Trans (Institute Ind Eng) 2012;44(4). https://doi.org/10.1080/0740817X.2011.564603.
[94] Nakazawa T, Kulkarni DV. Wafer map defect pattern classification and image retrieval using convolutional neural network. IEEE Trans Semicond Manuf 2018;31(2). https://doi.org/10.1109/TSM.2018.2795466.
[95] Kyeong K, Kim H. Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks. IEEE Trans Semicond Manuf 2018;31(3). https://doi.org/10.1109/TSM.2018.2841416.
[96] Wang J, Xu C, Dai L, Zhang J, Zhong R. An unequal deep learning approach for 3D point cloud segmentation. IEEE Trans Ind Informatics 2020;3203(c):1–11. https://doi.org/10.1109/TII.2020.3044106.
[97] Wang J, Xu C, Yang Z, Zhang J, Li X. Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition. IEEE Trans Semicond Manuf 2020;33(4):587–96. https://doi.org/10.1109/TSM.2020.3020985.
[98] Fernandez-Gauna B, Ansoategui I, Etxeberria-Agiriano I, Graña M. Reinforcement learning of ball screw feed drive controllers. Eng Appl Artif Intell 2014;30. https://doi.org/10.1016/j.engappai.2014.01.015.
[99] Liu T, Bao J, Wang J, Zhang Y. A hybrid CNN¯LSTM algorithm for online defect recognition of CO_2welding. Sensors (Basel) 2018;18(12). https://doi.org/10.3390/s18124369.
[100] Chien CF, Hsu CY, Hsiao CW. Manufacturing intelligence to forecast and reduce semiconductor cycle time. J Intell Manuf 2012;23(6). https://doi.org/10.1007/s10845-011-0572-y.
[101] Chen T. A systematic cycle time reduction procedure for enhancing the competitiveness and sustainability of a semiconductor manufacturer. Sustain 2013;5(11). https://doi.org/10.3390/su5114637.
[102] Liu Y, Hu X, Zhang W. Remaining useful life prediction based on health index similarity. Reliab Eng Syst Saf 2019;185(February):502–10. https://doi.org/10.1016/j.ress.2019.02.002.
[103] Sinha JK, Elbhbah K. A future possibility of vibration based condition monitoring of rotating machines. Mech Syst Signal Process 2013;34(1–2). https://doi.org/10.1016/j.ymssp.2012.07.001.
[104] Lei Y, Liu Z, Wu X, Li N, Chen W, Lin J. Health condition identification of multi-stage planetary gearboxes using a mRVM-based method. Mech Syst Signal Process 2015;60. https://doi.org/10.1016/j.ymssp.2015.01.014.
[105] Zhang XZ, Bi CX, Bin Zhang Y, Geng L. Real-time nearfield acoustic holography for reconstructing the instantaneous surface normal velocity. Mech Syst Signal Process 2015;52–53(1). https://doi.org/10.1016/j.ymssp.2014.06.009.
[106] Caesarendra W, Kosasih B, Tieu AK, Zhu H, Moodie CAS, Zhu Q. Acoustic emission-based condition monitoring methods: review and application for low speed slew bearing. Mech Syst Signal Process 2016;72–73. https://doi.org/10.1016/j.ymssp.2015.10.020.
[107] Wen L, Gao L, Li X, Zeng B. Convolutional neural network with automatic learning rate scheduler for fault classification. IEEE Trans Instrum Meas 2021;70. https://doi.org/10.1109/TIM.2020.3048792.
[108] Zheng H, Wang R, Yin J, Li Y, Lu H, Xu M. A new intelligent fault identification method based on transfer locality preserving projection for actual diagnosis scenario of rotating machinery. Mech Syst Signal Process 2020;135. https://doi.org/10.1016/j.ymssp.2019.106344.
[109] Wen L, Li X, Gao L. A new reinforcement learning based learning rate scheduler for convolutional neural network in fault classification. IEEE Trans Ind Electron 2020. https://doi.org/10.1109/TIE.2020.3044808.
[110] Yang B, Lei Y, Jia F, Xing S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mech Syst

Signal Process 2019;122:692–706. https://doi.org/10.1016/j.ymssp.2018.12.051.
[111] Ren S, Zhang Y, Liu Y, Sakao T, Huisingh D, Almeida CMVB. A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: a framework, challenges and future research directions. J Clean Prod 2019;210:1343–65. https://doi.org/10.1016/j.jclepro.2018.11.025.
[112] Xu C, Wang J, Zhang J, Li X. Anomaly detection of power consumption in yarn spinning using transfer learning. Comput Ind Eng 2021;152. https://doi.org/10.1016/j.cie.2020.107015.
[113] Gilmer J, et al. The relationship between High-dimensional geometry and adversarial examples. 6th Int. Conf. Learn. Represent. ICLR 2018 - Work. Track Proc. 2018.
[114] Su J, Vargas DV, Sakurai K. One pixel attack for fooling deep neural networks. IEEE Trans Evol Comput 2019;23(5):828–41. https://doi.org/10.1109/TEVC.2019.2890858.
[115] Marcus G. Deep Learning: A Critical Appraisal [Online]. Available: habr.com/post/371179/%0A. 2018. p. 1–24. https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf.
[116] Zhang Q, Zhu S-C. Visual interpretability for deep learning: a survey [Online]. Available:. 2018. http://arxiv.org/abs/1802.00614.
[117] David Bau AT, Zhou Bolei, Khosla Aditya, Oliva Aude. Network dissection: quantifying interpretability of deep visual representation seminar report. 2018.
[118] Ni CC, Lin YY, Luo F, Gao J. Community detection on networks with ricci flow. Sci Rep 2019;9(1):1–12. https://doi.org/10.1038/s41598-019-46380-9.
[119] Tallon PP. Corporate governance of big data: perspectives on value, risk, and cost. Computer (Long Beach Calif) 2013;46(6):32–8. https://doi.org/10.1109/MC.2013.155.
[120] Usman M, Jolfaei A, Jan MA. RaSEC: an intelligent framework for reliable and secure multilevel edge computing in industrial environments. IEEE Trans Ind Appl 2020;56(4):4543–51. https://doi.org/10.1109/TIA.2020.2975488.
[121] Wang J, Zheng P, Lv Y, Bao J, Zhang J. Fog-IBDIS: industrial big data integration and sharing with fog computing for manufacturing systems. Engineering 2019;5(4):662–70. https://doi.org/10.1016/j.eng.2018.12.013.
[122] Leng J, et al. Blockchain-secured smart manufacturing in industry 4.0: a survey. IEEE Trans Syst Man Cybern Syst 2021;51(1). https://doi.org/10.1109/TSMC.2020.3040789.
[123] Babiceanu RF, Seker R. Big Data and virtualization for manufacturing cyber-physical systems: a survey of the current status and future outlook. Comput Ind 2016;81:128–37. https://doi.org/10.1016/j.compind.2016.02.004.