IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 6, 2017, accepted January 6, 2018, date of publication January 15, 2018, date of current version February 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2793265

Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison

QINGLIN QI AND FEI TAO, (Senior Member, IEEE)
School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China
Corresponding author: Fei Tao (ftao@buaa.edu.cn)

This work was supported in part by the Beijing Nova Program in China under Grant Z161100004916063 and in part by the National Natural Science Foundation of China under Grant 51475032 and Grant 51522501.

ABSTRACT With the advances in new-generation information technologies, especially big data and digital twin, smart manufacturing is becoming the focus of global manufacturing transformation and upgrading. Intelligence comes from data. Integrated analysis for the manufacturing big data is beneficial to all aspects of manufacturing. Besides, the digital twin paves a way for the cyber-physical integration of manufacturing, which is an important bottleneck to achieve smart manufacturing. In this paper, the big data and digital twin in manufacturing are reviewed, including their concept as well as their applications in product design, production planning, manufacturing, and predictive maintenance. On this basis, the similarities and differences between big data and digital twin are compared from the general and data perspectives. Since the big data and digital twin can be complementary, how they can be integrated to promote smart manufacturing are discussed.

INDEX TERMS Big data, digital twin, smart manufacturing, comprehensive comparison, convergence.

I. INTRODUCTION

With the advances of the Internet, Internet of Things (IoT), big data, cloud computing, artificial intelligence (AI) and other new generation information technologies (New IT) [1], it brought valuable opportunities to many industries. As shown in FIGURE 1, more and more things are connected to the Internet. Gartner predicted that more than 20 billion devices (most from the manufacturing industry) would be connected to the IoT by 2020 [2]. As a result, a large volume of various data are generated, which would be over 40 zettabytes (ZB) by 2020 [3], including structured, semi-structured and unstructured data. With powerful storage and computing power of cloud computing, big data analysis models and algorithms are run to organize, analyze, and mine these raw data [4], [5], to obtain valuable knowledge. Meanwhile, AI with self-learning ability become more and more intelligent through data analytics.

In manufacturing, the big data involve a large volume of structured, semi-structured and unstructured data generated from the product lifecycle. The increasing digitalization of manufacturing is opening up opportunities for smart manufacturing [6]. The manufacturing data are collected real-time and automatically by IoT [7]. Through big data analysis based on cloud computing, manufacturers could find the bottlenecks of manufacturing processes, realize the causes and impacts of the problems, and find the solutions. So that the manufacturing processes are improved to enhance the manufacturing efficiency, making manufacturing more and more lean, and competitive. All the valuable information from manufacturing big data is feedback to product design, manufacturing, MRO (Maintenance, Repair & Overhaul), etc. It can help manufacturing achieve the change to smart manufacturing.

In addition, the interaction and convergence between the physical world and the cyber world of manufacturing is getting more and more attention. The digital twin paves a way to cyber-physical integration. Digital twin is to create the virtual models for physical objects in the digital way to simulate their behaviors [8]. The virtual models could understand the state of the physical entities through sensing data, so as to predict, estimate, and analyze the dynamic changes. While the physical objects would respond to the changes according to the optimized scheme from simulation [6]. Through the cyber-physical closed loop, digital twin could achieve the optimization of the whole manufacturing process [9].

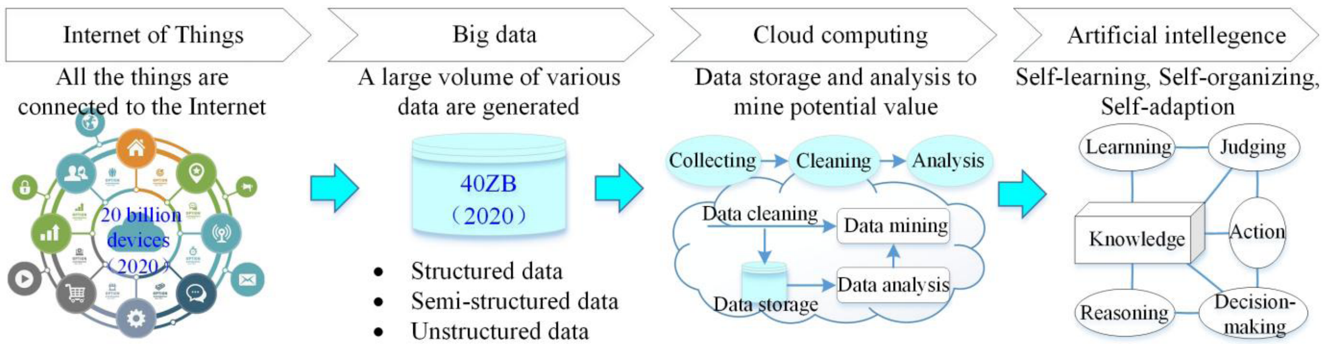

FIGURE 1. New IT and their applications.

In conclusion, collecting and analyzing a large volume of manufacturing data to find laws and knowledge, has become the key of smart manufacturing. Meanwhile, the digital twin breaks the barriers between the physical world and the cyber world of manufacturing. However, whether are there any similarities and differences between the big data and digital twin? What are their respective advantages? Is it possible to structure a bridge, which could bring big data and digital twin together? How to merge them together? These questions are all worth thinking deeply and being explored. Therefore, this paper reviews the concepts of big data and digital twin, as well as their applications in manufacturing. On this basis, they are compared from different aspects.

The main contributions of this paper include:

(1) The concepts of big data and digital twin are reviewed. And the data sources, data processing and data applications of big data in manufacturing are discussed, as well as the applications of digital twin in manufacturing.

(2) The similarities and differences between big data and digital twin in manufacturing, are compared from different aspects, including the general and data perspectives, as well as their respective advantages are discussed.

(3) The complementarity of digital twin and big data is discussed. And how the digital twin, big data and services are joined up to promote smart manufacturing, is illustrated.

The rest of this paper is organized as follows. The concepts of big data and the applications are reviewed in Section II, followed by the digital twin in Section III. The comparison of digital twin and big data in manufacturing is presented in Section IV. In Section V, the fusion of them in manufacturing is discussed. Finally, conclusions are drawn in Section VI.

II. BIG DATA IN MANUFACTURING

In the era of Internet of Everything (IOE), more and more physical devices are connected to the Internet, so that a large volume of data are collected by the RFID, sensors, gateways, etc., and transmitted through the IoT. Besides, mobile Internet, social networking, e-commerce, etc., greatly expand the applications of the Internet. As a result, various data are rapidly booming. In various industries, decision making are increasingly based on data and analysis, rather than experience. Data is becoming the important assets of human society, and big data era has come [10].

A. THE CONCEPT OF BIG DATA

There is no doubt that big data is becoming more and more important. However, there still are no unified opinions on the definition of big data. **In general**, big data is to describe a large amount of structured, semi-structured and unstructured data created by data sources, which would need too much time and money to be stored and analyzed to obtain huge value. Therefore, **for the data itself**, big data refers to the massive data that could not be collected, stored, managed, shared, analyzed, and computed by regular data tools within a tolerable time [11]. **For the users of data**, they pay more attention to the value of data rather than the enormous quantity [12]. Thus, big data is also interpreted as the ability to quickly acquire the hidden value and information from various and large amount of data. It goes beyond the general processing capabilities of users.

Besides, big data can also be defined by the following characteristics, which are Volume, Variety, Velocity, and Value, i.e., 4Vs [13]. With regards to Volume, it refers to that the data scale is very large, ranging from several PB (1000TB) to ZB (a billion TB) [14]. As for Variety, it means that the size, content, format, and applications of the data are diversified. For instance, the data include structured (e.g., digit, symbols, and tables), and semi-structured data (e.g., trees, graphs, and XML documents), and unstructured data (e.g., logs, audios, videos, documents, and images) [12]. Velocity means that the data generation is rapid, and the data processing requires high timeliness. In the face of massive data, velocity is the life of enterprises. For value, the significance of big data is not the great volume, but rather the huge value. How to extract the value from massive data through powerful algorithms, is the key to improve competitiveness. Furthermore, the characteristics of big data are extended to 10Vs, i.e., *Volume, Variety, Velocity, Value, Veracity, Vision, Volatility, Verification, Validation, and Variability* [15].

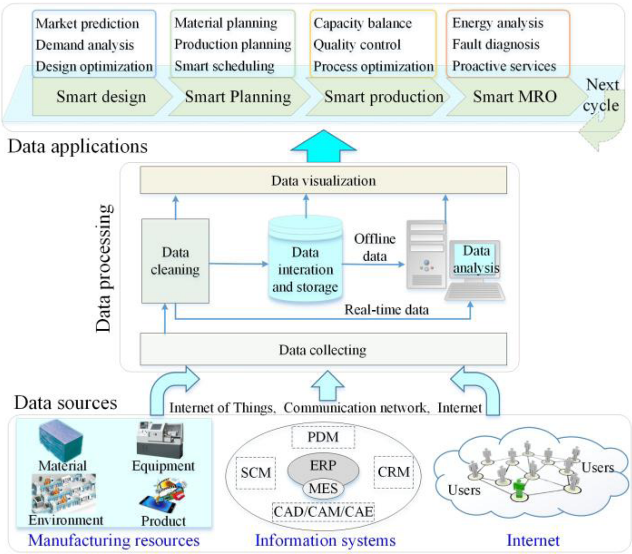

FIGURE 2. The sources, processing and applications of big data in manufacturing.

B. BIG DATA DRIVEN SMART MANUFACTURING

1) THE DATA SOURCES IN MANUFACTURING

In manufacturing, big data refer to the data generated from the product lifecycle, such as design, manufacturing, MRO, etc. [16], which are also featured with 4Vs, i.e., high volume (huge quantities of data), variety (the data itself comes in different forms and is generated by diverse sources), velocity (the data is generated and renewed at very high speed), and value (huge value hidden in the data). Manufacturing data are generally from the following aspects (see in FIGURE 2):

(1) Manufacturing resources data, including a) equipment data collected from smart factories by the Industrial IoT technologies, with respect to the real-time performance, operating condition, etc.; b) material and product data collected from themselves and service systems, such as performance, inventory, context of use, etc.; c) environmental data (e.g., temperature, humidity, air quality etc.), and so forth.

(2) Management data from manufacturing information systems (e.g., MES, ERP, CRM, SCM, and PDM) and computer aided systems (e.g., CAD, CAE, and CAM). Such data include, for example, designing scheme, order dispatch, material distribution, production planning, marketing and sales, service management, finance, and so forth.

(3) Internet data, including a) user data collected from the ecommerce platforms (e.g., Amazon, Walmart, and Taobao) and social networking platforms (e.g., Twitter, Facebook, LinkedIn, and YouTube), such as user comments, preference, and behaviors, etc.; b) public data from open websites (e.g., governments and public service websites), and so on.

2) THE DATA PROCESSING

However, only raw data are barely useful. It must be processed through several steps to extract the value (see in FIGURE 2). Firstly, the data are collected by various measures, such as IoT (e.g., smart sensors, and RFID) [17], [18], API (Application Programming Interface), SDK (software development kit), web crawler, etc. Because of the characteristics of multi-source, heterogeneous, multi-scale, high noise and others, manufacturing data need to be cleaned before further being processed [19]. Then, the clean data are integrated and stored for the exchange and sharing of manufacturing data at all levels. Next, based on cloud computing, the real-time data or off-line data are analyzed and mined, through the advanced analysis methods and tools, such as machine learning, forecasting models, etc. [20]–[22]. The valuable knowledge is extracted from the large number of dynamic and fuzzy data, enabling manufacturers to deepen their understandings of various stages of product lifecycle. Therefore, manufacturers will make more rational, responsive, and informed decisions and enhance their competitiveness.

3) THE APPLICATIONS OF BIG DATA IN MANUFACTURING

Big data has an unprecedented impact on the manufacturing, which involve the following aspects, as shown in FIGURE 2.

(1) In the big data era, product design is shifted from inspiration and experience-based design to data and analysis-driven design. Through the analysis for big data about user behaviors and market trend, designers could accurately quantify customer demands, to translate customer voices to product features and quality requirements [23]. As a result of big data analysis, the product design is significantly accelerated and optimized.

(2) Before manufacturing begins, smart production planning is conduct in consideration of manufacturing resources data. Based on the relationship of the global data (e.g., available resources and capacities information, material data, technological parameters, and constraints), the global and optimized planning program could be rapidly generated, improving the planning speed and accuracy.

(3) In the process of manufacturing, the real-time data enable the manufacturing process monitoring, so that the manufactures could keep abreast of the changes to develop the optimal operational control strategies [24]. Besides, in virtue of big data analysis, the product quality control and improvement are embedded into every step from raw material to finished product. For example, the early warning of quality defects, and rapid diagnosis of root causes of malfunctions, could be accomplished in real time to guarantee the high quality.

(4) Last, in virtue of the mighty predictive ability [25], big data changes the traditional and passive MRO mode. Through collecting and analyzing massive data from smart devices or products, it takes the initiative to carry

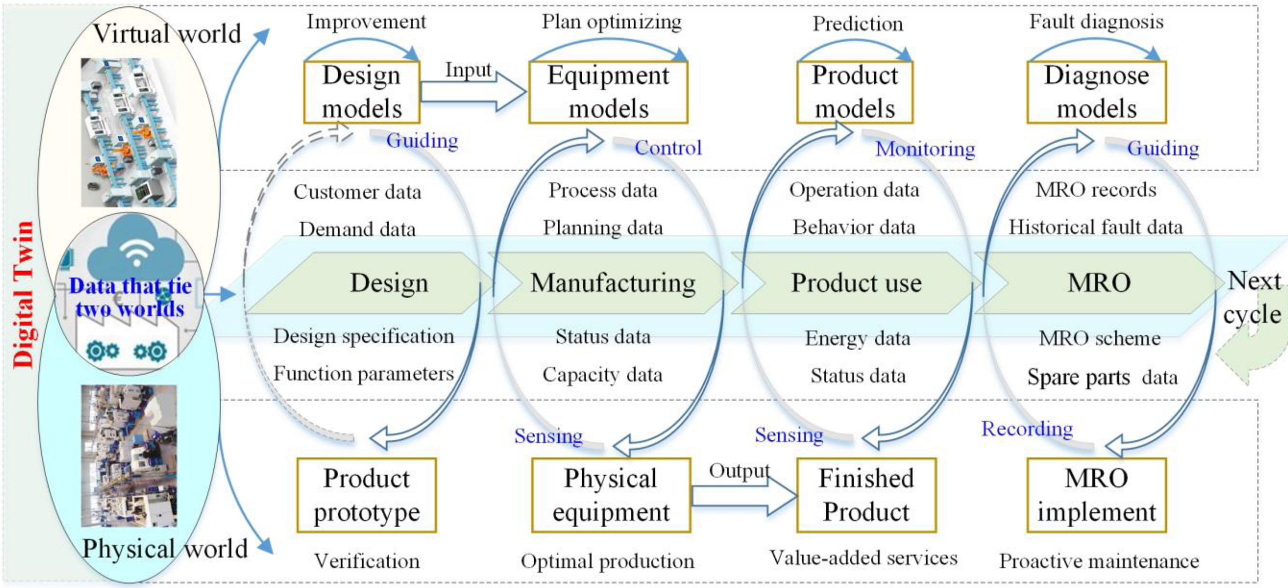

FIGURE 3. Digital twin in manufacturing.

out the devices or products health monitoring, fault diagnosis [26], and operation process optimization, etc., for active preventive MRO [27].

C. RECAPITULAION

From the above, it may get out such conclusions: First of all, the data is invaluable asset, which enable smart manufacturing. Secondly, the strategic significance of big data is not to master massive data, but rather to acquire the value with specific meaning through specialized processing. Thirdly, the value of converged data is far greater than the single type of data.

III. DIGITAL TWIN IN MANUFACTURING

Along with the rise of New IT, the scope, degree, functions of the cyber world of manufacturing, and the integration with the physical world, have been being strengthened [28]. Digital twin paves a way for cyber-physical integration. It serves as a bridge between the physical world and the cyber world, providing the manufacturing enterprises with a new way to carry out smart production and precision management.

A. THE CONCEPT OF DIGITAL TWIN

The concept of digital twin was firstly presented by Grieves [29]. In general, virtual models of physical objects are created in a digital way to simulate their behaviors in real-world environments [8]. Therefore, the digital twin is composed of three components, which are the physical entities in the physical world, the virtual models in the virtual world, and the connected data that tie the two worlds (see in Fig. 2) [29]. Digital twin reflects two-way dynamic mapping of physical objects and virtual models [28]. Specifically, it is the virtualization of physical entities. The physical operation process is judged, analyzed, predicted and optimized in virtual means. Corresponding, it is the materialization of virtual process. After the simulation and optimization of product design, manufacturing and maintenance process, it guides the physical process to perform the optimized solution [9]. In the process of interaction between virtual and reality, the integration of data is an inevitable trend. The data from physical world are transmitted to the virtual models through the sensors to complete the simulation, validation and dynamic adjustment. And the simulation data are fed back to the physical world to respond to the changes, improve the operation, and increase the value. Only on the basis of the converged data environment, cross analysis is possible.

B. THE APPLICATIONS OF DIGITAL TWIN IN MANUFACTURING

As shown in FIGURE 3, digital twin integrates all manufacturing processes, which can achieve the closed loop and optimization of the product design, manufacturing, and smart MRO, etc. [9]

1) DIGITAL TWIN BASED PRODUCT DESIGN

In the design phase, it involves back-and-forth interactions between the expected, interpreted, and physical worlds [30]. Based on digital twin, the digital representation (i.e., virtual models) of the physical product is created in the interpreted world (i.e., virtual world). The virtual models reflect both the expectations in the designer's mind, and the practical constraints in physical world. Digital twin enables the iterative optimization of design scheme to guide the designers to iteratively adjust their expectations and improve the design models, achieving personalized product design. In addition, digital twin driven virtual verification can quickly and easily forecast and verify product functions, behavior, structures and manufacturability, etc. [31]. Taking advantage of digital twin, it can accurately find the defect of design in virtual world and take rapid changes, which make the improvement of the design, avoiding tedious verification and testing.

2) SMART MANUFACTURING IN DIGITAL TWIN WORKSHOP/FACTORY

Next, the proven product design is input into the smart workshop or factory to be manufactured. From the input of raw material to the output of finished products, the whole manufacturing process is managed and optimized through digital twin [6]. The virtual workshop or factory include the geometrical and physical models of operators, material, equipment, tools, environment, etc., as well as the behaviors, rules, dynamics models and others [32]. Before they commit to manufacturing the products, the manufacturing resources and capacities are allocated, and production plan is devised to predefine the manufacturing process. The virtual workshop or factory simulate and evaluate the different manufacturing strategies and planning until a satisfactory planning is confirmed. In the actual manufacturing execution stage, the real-time monitoring and adjustment of manufacturing process are realized through virtual-physical interaction and iteration. The virtual models update themselves based on the data from the physical world, to keep abreast of the changes. And the problems are rapidly found out and the optimal solution is developed, through simulation in virtual world. According to simulation in virtual workshop or factory, the manufacturing process is adjusted to achieve optimal manufacturing (e.g., accuracy, stability, high efficiency and product quality).

3) PRODUCT DIGITAL TWIN FOR USAGE MONITORING

The virtual model of product is created to establish the product digital twin. The product digital twin would always keep in company with the product to provide the value-added services [33]. Firstly, the product in use is monitored in real time, as the product digital twin continually records the product usage status data, use environment data, operating parameters, etc. Consequently, users can keep abreast of the latest state of the product. Secondly, the virtual model can simulate the operation conditions of product in different environments. As a result, it can confirm what effects the different environmental parameters and operation behaviors would have on the health, lifetime, and performance, etc., so as to control the status and behaviors of physical product (e.g., change the operating parameters). Thirdly, based on the real-time data from physical product and historical data, the product digital twin is able to accurately predict the product remaining life, faults, etc. [34].

4) DIGITAL TWIN AS ENABLER FOR SMART MRO

Based on the prediction for health condition, remaining life, and faults, the proactive maintenance is carried out to avoid the sudden downtime. Furthermore, when a fault occurs, with the ultra-high-fidelity virtual model of the product, the fault would be visually diagnosed and analyzed [35], so that the position of faulty part and the root cause of fault are displayed to users and servicemen. Thereby, the MRO strategies (e.g., disassembly sequence, spare parts, and required tools) are developed to recovery the product. However, before starting the actual MRO (both proactive and passive), the walkthrough about MRO strategies would be executed in the virtual world based on virtual reality and augmented reality. As the mechanical structure of the parts and the coupling between each other are faithfully reflected by the virtual models, it can identify whether the MRO strategies are effective, executable and optimal. Once the MRO strategies are determined, they will be executed to recovery the product. Last, the data from the different stage of product lifecycle are accumulated and inherited to contribute to the innovation of the next generation product.

C. RECAPITULAION

In conclusion, the digital twin brings together the data from all aspects of product lifecycle, laying the data foundation for innovative product design and the quality traceability. Besides, the digital twin promotes the efficient synergies between the different stage of product lifecycle, achieving the iterative optimization. Furthermore, the digital twin shortens the product development cycle, improves the manufacturing efficiency and ensures the accuracy, stability, and quality.

IV. 360-DEGREE COMPARISON OF DIGITAL TWIN AND BIG DATA IN MANUFACTURING

Both big data and digital twin has attracted a wide spread attention, and are considered as the key to smart manufacturing. In order to evaluate the similarities and differences in big data and digital twin, the different aspects are compared as exemplified in Table 1.

A. THE COMPARISON FROM THE GENERAL PERSPECTIVE

1) THE SIMILARITIES IN DIFFERENCES

First of all, the origin of big data is from the exponential growth of the data, which is the result of advance of information technologies. While digital twin is to respond to the desire for interaction and integration between the physical and cyber worlds, which is inseparable from the rapid popularization and application of information technologies [34]. Therefore, although the initial focus of big data and digital twin is not the same, the big background of both them is roughly same, namely the advance and wide applications of New IT.

All of the functions of big data are from data processing. Through big data analysis, it can identify the behavior features and patterns, and have an insight into the trends, to help users make decisions. The prediction or optimization abilities of big data need for training data set, or comparing with historical results. While the functions of digital twin rely on the simulation and evolution of the virtual models. Due to the

TABLE 1. The comparison between digital twin and big data in manufacturing.

Items	Big data	Digital Twin
Back-ground	The rapid development and wide applications of New IT and the exponential growth of the data	The rapid development of New IT and the desire for cyber-physical integration
Concep t	Including the data and the processing; Focusing the large volume and value	Including the physical and virtual worlds and data that tie two worlds; Focusing the virtual-real dual-reflection
Func-tions	Mining behavior features and patterns; Insight into the trends; Data visualization; Predicting and analyzing the problems; Aiding decision making; Optimizing and improving the process	Virtual verification; Simulation running; Ultra-high-fidelity real-time monitoring; Predicting and diagnosing the problems; Optimizing and improving the process
Applica -tions	Product lifecycle from design to MRO, etc.	Product lifecycle from design to MRO, etc.
Effects	Improve efficiency, customer satisfaction and precision of management; Extend life of product and equipment; Reduce the cost; Promote smart manufacturing	Improve efficiency, customer satisfaction and precision of management; Extend life of product and equipment; Reduce the cost and development cycle; Promote smart manufacturing
Key techno-logies	IoT; Cloud computing; Fog computing; Data cleaning; Data mining; Machine learning, etc.	IoT; Data analysis; Virtual reality; Augmented reality; CPS; Simulation, etc.
Data sources	The data from the physical entities, information systems and Internet in every stage of the product lifecycle	The data from the physical entities, virtual models, and their fusion in every stage of the product lifecycle
Data volume	Large, ranging from PB to EB, even ZB	No specific quantity
Data feature s	Structured, semi-structured and unstructured data	Structured, semi-structured and unstructured data
Multi-source correla -tion	Focus on data attributes and highlight the relationships between features	Focus on the consistency of multi-source data and their evolution and integration
Data acquisi -tion	Sensors, RFID, and other sensing devices; SDK, API; Web crawler, etc.	Sensors, RFID, and other sensing devices; Model data interfaces, etc.
Data process -ing	Through big data processing tools, algorithms platforms, etc.	No specific methods
Data fusion	Various objects data fusion in single phases of the product lifecycle	The all elements, whole process, whole business data fusion in the entire product lifecycle
Visuali-zation	Table, chart, graphs, and file printing, etc.	Image, video, virtual and augmented reality, etc.
Result verifica -tion	Through the physical execution process, or the simulation from the third party; Relatively slow	Through its own virtual simulation and evolution functions to execute pre-verification; Relatively advance

ultra-high-fidelity virtual models, the digital twin can simulate the entire operation process independently, according to the actual operating rules of the physical world. Compared with big data, the digital twin can visually run and verify the manufacturing process in the virtual world. However, from their functions, although the implementation methods have their own characteristics, both of them share a majority of same functions. For example, both of them could predict and diagnose the problems, as well as real-time monitor, optimize and improve the manufacturing process.

In addition, the same functions lead to the similar fruit effects. Both of them improve production efficiency, customer satisfaction and precision of management, as well as reduce the cost in their own way, promoting smart manufacturing. Moreover, they also extend the life of product and equipment, reduce development cycle.

In conclusion, although big data and digital twin differ with each other in detail, they are consistent with the overall direction of the background, functions, and effects. Therefore, they have the foundation of cooperation with each other.

2) THE DIFFERENCES AND COMPLEMENTARITY UNDER SIMILARITIES

With respect to the concept, the contents of big data are relatively simple, emphasizing the large scale and value hidden in the data. In order to extract the value from massive data, big data uses advanced tools and algorithms different from traditional ones [13]. Whereas, the digital twin is composed of three components, of which data is an integral part. The digital twin is more concerned with the virtual-real dual-reflection. From the concept, although both of them involve data, big data is more professional and efficient than digital twin in data. Therefore, big data can serve for digital twin.

In the term of applications, both of them are applied in every stage of the product lifecycle from design to MRO, etc. However, compared with digital twin, there are barriers between the various phases of lifecycle in the application of big data in manufacturing. As the designer, manufacturer, and serviceman, etc., may not be from the same company, the data from a certain phase sometimes may only be used in its own phase. Taking the interests and data sharing security into account, it does not fully realize the continuous flow of data in the product lifecycle. In contrast to it, the digital twin can collect, record, accumulate, and comprehensively process all the data from product design until retirement. It not only be conducive to the design manufacturing, use and MRO of the product, but also contributes to the next generation product. Therefore, the digital twin can make up the deficiency of large data to break the barriers in the product lifecycle.

As well, big data and digital twin share some key technologies, such IoT. However, big data focuses more on the technologies about data (e.g., cloud computing, data cleaning, data mining, and machine learning), while digital twin care more about the technologies about cyber-physical integration (e.g., simulation, virtual reality, augmented reality, and CPS). The combination of their key technologies will be more effective for their application in the product lifecycle.

In conclusion, despite the different emphasis, both of them have their own strengths. Moreover, their advantages are complementary, to make up for their own deficiencies. Therefore, they have the bonding point to cooperate with each other.

B. THE COMPARISON FROM THE DATA PERSPECTIVE

The data are common content to both of big data and digital twin. Their respective advantages in terms of data are analyzed as following.

1) THE ADVANTAGES OF BIG DATA OVER DIGITAL TWIN

First of all, the large scale is inherent in the concept of big data. The data volume of big data is ranging from PB to EB, even ZB. The abundant data means rich information and knowledge. In contrast to it, the digital twin does not indicate the specific quantity of data. In general, as an important part of the digital twins, the data should not be insufficient. However, by comparison, the data that are collected, stored, managed, analyzed, and computed in big data, are certainly more than digital twin. Therefore, big data is more suitable for extracting more knowledge from larger volume of data.

In the term of multisource correlation, big data focuses on the data attributes and highlight the relationships between features. Big data is to find the correlation relationship and knowledge through mining behavior features and patterns. The digital twin focuses on the consistency of multi-source data. It is to simulation and deduce the manufacturing process. The inconsistency of data would lead to conflict. Therefore, big data is more suitable for more varieties of data.

Besides, because the massive data cannot be processed by regular data tools within a tolerable time, big data processing has its advanced tools, algorithms, platforms, etc. While the data processing methods is not specified in digital twin. As the data is professionally processed, therefore, the velocity of big data in data processing is more efficient than digital twin.

In conclusion, big data is more professional and efficient than digital twin in terms of volume, value, variety, and velocity of data, which are consistent with the characteristics of big data. It further deepens the role that big data can serve for digital twin.

2) THE ADVANTAGES OF DIGITAL TWIN OVER BIG DATA

As the digital twin is composed of three components, the data sources of digital twin is different from big data. The data of big data are from the physical entities, information systems and Internet, which are all generated by activities in physical world. The data in digital twin are not only from the physical world, but also from the virtual models. In addition, some data are derived from the fusion operation for the data from the two worlds, such as synthesis, statistics, association, clustering, evolution, regression and generalization. According to the different data sources, the data acquisition tools are also different. Although both of them share the same tools for data from physical world (e.g., sensors, and RFID), digital twin needs model data interfaces to collect data from virtual world, which is not needed in big data in manufacturing. Therefore, digital twin has more comprehensive data to be used.

As well, the data fusion of big data is the fusion of various objects data in single phase of the product lifecycle, due to the barriers between different phases. Whereas it is the all elements, whole process, whole business data fusion in digital twin. The digital twin can achieve the data sharing and integration between different phases of product lifecycle. Therefore, digital twin extends the range of application of manufacturing data, and avoid the duplicate and waste.

With respect to the visualization, big data prefer to use the two-dimensional and static tools, such as table, chart, graphs, and file printing, etc. Because of the virtual models, the visualization in digital twin is more visual, which is mostly three-dimensional and dynamic, such as image, video, virtual and augmented reality, etc. Furthermore, because the object of big data is just data, in order to verify analysis results, it must be through the physical execution process, or the simulation from the third party, which is relatively slow. Whereas with its own virtual simulation function, the digital twins can complete the pre-verification of the results in virtual world. Therefore, digital twin is more advance and convenient about the visualization and result verification.

In general, from the data perspective, big data is more powerful than digital twins in data, while digital twin is better than big data in applications. Therefore, it is a better option to develop digital twin with big data.

V. THE FUSION OF DIGITAL TWIN AND BIG DATA IN MANUFACTURING

A. THE COMPLEMENTARITY BETWEEN BIG DATA AND DIGITAL TWIN

As the manufacturing process becomes more and more complex, it is difficult to quickly identify the problems arising in the manufacturing process by traditional way. These visible and invisible problems in smart manufacturing, can be reflected by the data. Big data brings more efficiency, sharper insight and more intelligence to manufacturing. Besides, digital twin help realize the correlation and dynamic adjustment between manufacturing planning and implementation, as well as promote faults prediction, diagnosis and maintenance in digital way. Although there are many differences between digital twin and big data, they play the complementary roles in the manufacturing.

As shown in FIGURE 4, big data could be considered as an important part of digital twin. Without big Data, most of functions of digital twin would be the castle in the air. And without digital twin, the big data analysis and the actual manufacturing process would not be in parallel. The convergence of digital twin and big data can break the barriers between different phases of product lifecycle, and shortens the product development and verification cycle. Moreover, embracing the concept of Manufacturing-as-a-Service (MaaS), service-oriented smart manufacturing receives extensive attention. As an effective means, services enable numerous large-scale manufacturing collaboration. Therefore, digital twin, big data and services can be united to promote smart manufacturing.

B. THE PERSPECTIVE IN FUSION OF DIGITAL TWIN AND BIG DATA IN MANUFACTURING

In the design phase, product innovation relies on the accurate interpretation of market preferences and customer demands,

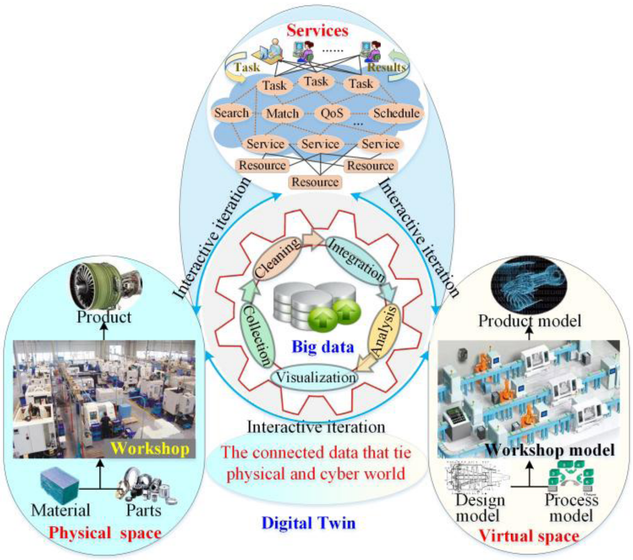

FIGURE 4. The fusion of digital twin, big data and services in manufacturing.

and the capacity to translate customer voices to product features and quality requirements. The big data enables designers to keep abreast of customer demands. Based on big data analysis, the function structure and components of product are designed in virtue of digital twin. With the different view of traditional methods, the product prototype does not need to be first manufactured to facilitate the manufacturers to assess the design quality and feasibility. Designers can create vivid simulation scenarios to predict the finished quality of designed product, and identify the design defects in virtual world. In this case, whether the design parts can be manufactured, whether the parts can be assembled, whether parts interfere with each other, and whether the design scheme meet the relevant specifications and functions requirements, can be quickly verified. If the design scheme could not pass the simulation testing, it must be redesigned in real-time. When the product is redesigned, the big data would be used to identify problems to improve design scheme. In the process of product design, the required tools and algorithms are used in the form of services.

In the manufacturing stage, first of all, hypernetwork based manufacturing resource services supply-demand matching and scheduling [36] are carried out to quickly find available resources. On this basis, all the required resources are integrated together for analysis and planning through big data analysis. The production plan is simulated, evaluated, and improved in virtual world. After acquiring the best production plan, it is delivered to the physical world to implement the actual production. Meanwhile, the real-time data is collected from the physical world, to drive the virtual models to monitor the manufacturing process, which are compared with plans. If there are differences, big data analysis is used to find out the reasons and develop solutions, such as adjusting the equipment or improving the plan. In the iterative interaction, it ensures that the production can be fully implemented in accordance with the optimal planning. Besides, once the design changes, the manufacturing process can be easily updated, including updating the bill of materials, processes, and assigning new resources. As a result, the convergence of digital twin, big data and service, enables the production planning optimizing and manufacturing process real-time adjustment.

In the daily operation and MRO of the product, the virtual models of physical products are synchronized with the real state of the product through the sensors. The operation status of the product, and the health status of the components, are grasped in real time. In addition to the sensors data, product digital twin also integrates historical data (e.g., maintenance records, and energy consumption records). Through big data analysis of the above data, product digital twin can continually predict the health state of the product, the remaining life of the product and the probability of faults. In virtue of big data analysis, product digital twin can also reveal the unknown problems by comparing the actual product response and anticipating product responses in specific scenario. Once the hidden problems are found, or faults occur, the maintenance programs are simulated and optimized in the virtual world, to facilitate the actual maintenance. Similarly, the resources and capabilities, tools and algorithms required in the daily operation and MRO stage are used in the form of services. Therefore, the product life and the maintenance efficiency are improved and maintenance time and cost are reduced.

Big data analysis is responsible for analyzing all the data required by smart manufacturing, while digital twin makes up the drawbacks, which are that big data could not simulate and synchronously visualize physical process. Therefore, the convergence of digital twin, big data and service, is of great significance to smart manufacturing.

VI. CONCLUSION

Both of digital twin and big data play important roles in promoting smart manufacturing. Digital twin enables manufacturers to manage the real-time and two-way mappings between physical object and digital representation, which paves the way for cyber-physical integration. In combination with the accurate analysis and prediction capabilities of big data, the digital twin driven smart manufacturing will be made more responsive and predictive and will be beneficial to more reasonable and precise manufacturing management in many aspects. Together they complement each other nicely to help the development of smart manufacturing.

However, the smart manufacturing is very complex, as well as its practical applications and dynamic evolution. This paper preliminarily investigated the roles of big data and digital twin in smart manufacturing. It still needs further researches to improve and enrich related works by considering more factors and practical situations. Some further works are pointed out and not limited to as follows:

(1) Considering the virtual-reality interaction environment of the digital twin, the efficient whole elements, whole

business and whole process data integration and fusion algorithm, models and platforms need future research.

(2) Considering the various demands of all parties involved in service collaboration, the research on the comprehensive utility balance of manufacturing services is of great significance to maximize the value of digital twin.

REFERENCES

[1] F. Tao and Q. Qi, "New IT driven service-oriented smart manufacturing: Framework and characteristics," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2723764.

[2] D. Mourtzis, E. Vlachou, and N. Milas, "Industrial big data as a result of IoT adoption in manufacturing," *Procedia CIRP*, vol. 55, pp. 290–295, Nov. 2016.

[3] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," *IDC iView: IDC Analyze Future*, vol. 2007, no. 2012, pp. 1–16, Dec. 2012.

[4] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, Jan. 2015.

[5] S. Yi, C. Li, and Q. Li, "A survey of fog computing: Concepts, applications and issues," in *Proc. Workshop Mobile Big Data*, Hangzhou, China, Jun. 2015, pp. 37–42.

[6] R. Rosen, G. V. Wichert, G. Lo, and K. D. Bettenhausen, "About the importance of autonomy and digital twins for the future of manufacturing," *IFAC-Papersonline*, vol. 48, no. 3, pp. 567–572, Aug. 2015.

[7] F. Tao, Y. Zuo, L. Xu, and L. Zhang, "IoT-based intelligent perception and access of manufacturing resource toward cloud manufacturing," *IEEE Trans. Ind. Inf.*, vol. 10, no. 2, pp. 1547–1557, Feb. 2014.

[8] J. Hochhalter *et al. Coupling Damage-Sensing Particles to the Digitial Twin Concept*. Accessed: Jan. 17, 2018. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=20140006408

[9] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui. (2017). "Digital twin-driven product design, manufacturing and service with big data," *Int. J. Adv. Manuf. Technol.*, Mar. 2017, doi: 10.1007/s00170-017-0233-1.

[10] J. Bughin, M. Chui, and J. Manyika, "Clouds, big data, and smart assets: Ten tech-enabled business trends to watch," *McKinsey Quart.*, vol. 56, no. 1, pp. 75–86, 2010.

[11] J. Manyika *et al.*, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. San Francisco, CA, USA: The McKinsey Global Institute. [Online]. Available: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

[12] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015.

[13] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.

[14] J. Gantz and D. Reinsel, "Extracting value from chaos," *IDC iview*, vol. 1142, no. 2011, pp. 1–12, 2011.

[15] R. F. Babiceanu and R. Seker, "Big data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook," *Comput. Ind.*, vol. 81, pp. 128–137, Sep. 2016.

[16] J. Li, F. Tao, Y. Cheng, and L. Zhao, "Big data in product lifecycle management," *Int. J. Adv. Manuf. Technol.*, vol. 81, nos. 1–4, pp. 667–684, Oct. 2015.

[17] Y. Zhang, G. Zhang, J. Wang, S. Sun, S. Si, and T. Yang, "Real-time information capturing and integration framework of the Internet of manufacturing things," *Int. J. Comput. Integr. Manuf.*, vol. 28, no. 8, pp. 811–822, Aug. 2015.

[18] A. Kumar, H. Kim, and G. P. Hancke, "Environmental monitoring systems: A review," *IEEE Sensors J.*, vol. 13, no. 4, pp. 1329–1339, Apr. 2013.

[19] S. Wang, L. I. Yuchen, N. Liu, and S. Wang, "Noisy-data-disposing algorithm of data clean on the attribute level," *Comput. Eng.*, vol. 31, no. 9, p. 86, 2005.

[20] C. Zhu, H. Wang, X. Liu, L. Shu, L. T. Yang, and V. C. Leung, "A novel sensory data processing framework to integrate sensor networks with mobile cloud," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1125–1136, Sep. 2016.

[21] K. W. Lin and D. J. Deng, "A novel parallel algorithm for frequent pattern mining with privacy preserved in cloud computing environments," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 6, no. 4, pp. 205–215, 2010.

[22] A. Siddiqa *et al.*, "A survey of big data management: Taxonomy and state-of-the-art," *J. Netw. Comput. Appl.*, vol. 71, pp. 151–166, Aug. 2016.

[23] Y. Liu, J. Jin, P. Ji, A. Harding, and R. Y. Fung, "Identifying helpful online reviews: A product designer's perspective," *Comput. Aided Des.*, vol. 45, no. 2, pp. 180–194, Feb. 2013.

[24] L. Wang, W. Shen, and S. Lang, "Wise-ShopFloor: A Web-based and sensor-driven e-shop floor," *J. Comput. Inf. Sci. Eng.*, vol. 4, no. 1, pp. 56–60, Mar. 2004.

[25] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renew. Energy*, vol. 48, pp. 110–116, Dec. 2012.

[26] Z. Huo, M. Mukherjee, L. Shu, Y. Chen, and Z. Zhou, "Cloud-based data-intensive framework towards fault diagnosis in large-scale petrochemical plants," in *Proc. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Paphos, Cyprus, 2016, pp. 1080–1085.

[27] J. Wan *et al.*, "A manufacturing big data solution for active preventive maintenance," *IEEE Trans. Ind. Inf.*, vol. 13, no. 4, pp. 2039–2047, Aug. 2017, doi: 10.1109/TII.2017.2670505.

[28] F. Tao and M. Zhang, "Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing," *IEEE Access*, vol. 5, pp. 20418–20427, Sep. 2017.

[29] M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication," White Paper, 2014. [Online]. Available: http://innovate.fit.edu/plm/documents/doc_mgr/912/1411.0_Digital_Twin_White_Paper_Dr_Grieves.pdf

[30] J. S. Gero and U. Kannengiesser, "The situated function–behaviour–structure framework," *Des. Stud.*, vol. 25, no. 4, pp. 373–391, Jul. 2004.

[31] B. Schleich, N. Anwer, L. Mathieu, and S. Wartzack, "Shaping the digital twin for design and production engineering," *CIRP-Ann. Manuf. Technol.*, vol. 66, no. 1, pp. 141–144, 2017.

[32] F. Tao, Y. Cheng, J. Cheng, M. Zhang, W. Xu, and Q. Qi, "Theories and technologies for cyber-physical fusion in digital twin shop-floor," (in Chinese), *Comput. Integr. Manuf. Syst.*, vol. 23, no. 8, pp. 1603–1611, Aug. 2017.

[33] C. Zhuang, J. Liu, H. Xiong, X. Ding, S. Liu, and G. Wang, "Connotation, architecture and trends of product digital twin," (in Chinese), *Comput. Integr. Manuf. Syst.*, vol. 23, no. 4, pp. 753–768, Apr. 2017.

[34] E. J. Tuegel, A. R. Ingraffea, T. G. Eason, and S. M. Spottswood, "Reengineering aircraft structural life prediction using a digital twin," *Int. J. Aerosp. Eng.*, vol. 11, Aug. 2011, Art. no. 154798.

[35] B. T. Gockel, A. W. Tudor, M. D. Brandyberry, R. C. Penmetsa, and E. J. Tuegel, "Challenges with structural life forecasting using realistic mission profiles," in *Proc. 53rd AIAA/ASME/ASCE/AHS/ASC Struct. Dyn. Mater. Conf. 20th AIAA/ASME/AHS Adaptive Struct. Conf. 14th AIAA*, Honolulu, HI, USA, 2012, p. 1813.

[36] F. Tao, J. Cheng, Y. Cheng, S. Gu, T. Zheng, and H. Yang, "SDMSim: A manufacturing service supply-demand matching simulator under cloud environment," *Robot. Comput. Integr. Manuf.*, vol. 45, pp. 34–46, Jun. 2017.

QINGLIN QI received the B.S. degree in computer science and technology from the Ocean University of China in 2012 and the M.S. degree in computer science and technology from Beihang University, China, in 2015, where he is currently pursuing the Ph.D. degree in control theory and control engineering.

His research interests include digital twin, manufacturing service, and smart manufacturing.

FEI TAO received the B.S. degree and the Ph.D. degree in mechanical engineering from the Wuhan University of Technology in 2003 and 2008, respectively. He is currently a Professor and the Vice Dean with the School of Automation Science and Electrical Engineering, Beihang University.

His research fields include service-oriented smart manufacturing, manufacturing service management, sustainable manufacturing, and digital twin-driven product design/manufacturing/service. He has authored four monographs and over 100 journal papers in these fields.

He is currently an Editor of the *International Journal of Service and Computing Oriented Manufacturing*.