International Journal of Information Technology Vol. 25 No. 2 2019

Brain Meets Brawn: Why Digital Twin and Knowledge Graph Needs Each Other In Industrial AI?

Huanhuan Chen

School of Computer Science and Technology, University of Science and Technology of China, Hefei, China

hchen@ustc.edu.cn

Abstract

The digital twin and knowledge graph are both pursuing the development of Industrial AI from different perspectives. The digit twin focused on what “brawn”: infrastructure, and tools for data collection, optimization and manipulation of the physical twin. In contrast, knowledge graph concentrated on “brain,” i.e., the development of concepts, methodologies, and algorithms for automatic knowledge extraction, management and utilization for the industrial dynamic environments. As these two techniques becomes mature and are applicable to real-world applications. Both are encountering the challenging problems in different aspects of scalability and applications. This process encourages an increasing interest to combine the two research areas. Motivated by this, we review the challenges of the two techniques, and propose some potential research problems when integrate the two.

I. Introduction

Industrial artificial intelligence (industrial AI), which refers to the application of AI techniques to industry, has received more attention in both the research and industry communities. Industrial machines and systems will become more intelligent, autonomous and even predictive as new techniques in industrial AI are integrated to improve the industrial machines, services and manufacturing processes with dynamic, new capabilities.

The digital twin is an emerging technique in industrial AI. The aim is to build a digital replica of actual physical assets, processes, systems and devices. The digital twin is an interface of the physical system. It could improve the digital twin by the related techniques, such as monitoring, optimization, and self-organization without interacting with the actual physical system [1].

The smooth operation of the digital twin relies on real-time data collection and processing. Most of the current techniques are data-driven. As the number of sensors increases in the digital twin, the volume of the collected data is huge. An alternative way is to explicitly incorporate the human knowledge/rules in the digital twin to facilitate the fast processing and to understand the dynamics of the system.

Knowledge engineering is one of the building blocks of AI, and it attempts to emulate the behavior of a human expert in a specific domain. The key research topic includes knowledge representation, extraction, fusion, utilization, etc. The knowledge graph (KG) is an essential technique to represent and utilize the knowledge.

The digital twin and knowledge graph are both pursuing the development of Industrial AI from different perspectives.

The digit twin focused on what we refer to as “brawn”: infrastructure, tools for data collection, optimization, and manipulation of the physical twin. In contrast, knowledge graph concentrated on “brains,” i.e., the development of concepts, methodologies, and algorithms for automatic knowledge extraction, management and utilization for the industrial dynamic environments. This research is motivated to incorporate the knowledge of human experts and to incorporate the knowledge into the data-driven learning models.

As these two techniques becomes mature and are applicable to real-world applications.

Both are encountering the challenging problems in different aspects of scalability and applications. This process encourages an increasing interest to combine the two research areas. Specifically, the current digital twin system is expensive in design and operation in terms of the time and effort, while knowledge graph is typically not incorporated with data-driven learning model that need to scale.

Nevertheless, the working areas of the both are becoming overlap, as digital twin seek to incorporate human knowledge to become more effective and flexible, and knowledge graph systems seek to be more data/system oriented.

II. Digital Twin

Digital Twin is the cornerstone for manufacturing process management. The “Digital Twin” was introduced in 2003 with the concept to create a virtual, digital equivalent to a physical product at University of Michigan Executive Course on Product Life-cycle Management (PLM) [2]. It is consisted of three virtual space, including physical products in real space, virtual products in virtual space, and the connections between the virtual and real products. Since digital twin was introduced, there has been tremendous increases in research of both the physical and virtual products. With the development of Industrial Internet of things, more real-time data is being collected for further improvement of digital twin.

2.1 Technologies

According to the work of Lu et.al [8], most of the existing research on digital twin is conceptual work by comparing the total number of publications on this topic, which indicates that it is still an early stage to develop practical digital twin applications.

The related research of digital twin includes the following aspects:

- The type of the digital twin, i.e., manufacturing devices, factory or production network. Most of digital twin applications are developed for single manufacturing devices. There are just a few applications for production networks indicating that the difficulty of Digital Twin grows with the scalability.
- The information model of the digital twin: the information models describe the data structure, semantics and logics in Digital Twin. There are some information models for manufacturing assets. But, it is obvious that the information model for a factory has not been fully exploited. This could be achieved by integrating the existing information models for manufacturing assets.

- The benefit(s) or purpose(s) of the digital twin application: The digital twin is mostly applied to the manufacturing assets monitoring for fault diagnosis and assets prediction for better management and process optimization. Most applications in digital twin function as decision-making support. There are very few work focusing on the feedback control from Digital Twin to a physical object.

In Digital Twin, the data collected from various sensors will be Big Data [3]. How to effectively process the Big Data gathered from the physical space is essential for developing a Digital Twin. The data processing, perdition and related AI techniques with the ability to address conflicts between single data records, the low-latency speed, and the noise robustness ability are required in a successful deployed Digital Twin.

III. Knowledge Graph

Knowledge engineering is one of the building blocks of AI, and it attempts to emulate the behavior of a human expert in a specific domain.

Knowledge engineering [4] is one of the AI techniques to acquire, manage and utilize the knowledge. It includes three fundamental scientific issues: knowledge acquisition, knowledge representation, and knowledge utilization/inference. The data collected for big data applications often comes from heterogeneous, autonomous information sources for complex and evolving relationships [5]. In such situations, knowledge does not rely only on domain expertise, but also fragmented knowledge pieces from multiple information sources.

3.1 Technologies

In many systems such as digital twin, storing all observed data is very expensive or infeasible for large scaled application, which leads to efficiently acquire useful knowledge from different local sources. However, real-time data stream usually comes from multiple, heterogeneous, autonomous sources with complex relationships, and some algorithms can typically acquire only pieces of knowledge from single sources [4]. In addition, those pieces of knowledge are usually fragmented with properties as uncertainty, incompleteness, and varying degrees of quality.

The construction and deployment of a KG system rely on carefully designed knowledge acquisition, representation, and inference. Knowledge engineering is also a cornerstone for various intelligent systems, which has the features of context awareness and adaptability.

The challenge of knowledge engineering lies on the lack of a uniform representation for knowledge. Knowledge graph (KG) is an effective representation for knowledge with the definition of entities and their relations. KG is able to process various types of data, such as structure, semi- structure, text data, and so on. Due to various data sources, the non-uniform data formats make it difficult to form a uniform logic of methodology to extract the useful knowledge. In some situations, we have to abstract some details of the data that could sacrifice the accuracy.

The inks in a knowledge graph are established by discovering the relationships among data objects. To extract real-time knowledge from digital twin, the corresponding knowledge graph should be updated accordingly. It is necessary to determine 1) an appropriate time interval of the updating, and 2) which entities or the relationship in the previous knowledge graph, should be modified, in the dynamical update process of a knowledge graph.

It is often required to construct a multiple-granularity KG to satisfy the requirements of real-world applications. In practices, we would need to adjust the granularity according to actual application scenarios.

IV. Brains and Brawn: Three potential Research Problems

As discussed before, the two research areas focused on different aspects in industrial AI. There are several research directions to combine digital twin with knowledge graph, including: a) to construct the (event) knowledge graph from the real-time data to accelerate the data processing or to assist the decision making in digital twin; and b) to incorporate the prior human knowledge into the model design, construction and optimization [9].

In the case of digital twin, the primary concern is to improve the efficiency and effectiveness of the system with prior knowledge. Thus, the effort has been devoted to how the knowledge graph is constructed via real-time data, how to use the knowledge graph to monitor, and improve digital twin with missing data presented.

In the case of knowledge graph, the effort has been devoted to the fundamental mechanisms, including knowledge representation and extraction from multiple sources; knowledge fusion from different sub-graph; and automated demand-driven knowledge services. KG provides a route to adaptively extract the knowledge from data, but it has not been fully employed in the systematical real-world applications. Therefore, the integration of the digital twin and KG is a reasonable combination of brain and brawn. In the following, we outline three potential areas (in no particular order) in which research is needed to realize an integrated digital twin and KG.

- **Explainable Digital Twin dynamics model**

  Currently, most of the traditional learning models fall into the category of the so-called data-centred approaches, which operate in the space of data measurements to learn discriminative models. However, this kind of approaches is usually a “black box”, which is not able to capture the underlying mechanical processes that generated the data. The knowledge graph based approaches operate in a “white-box” style, and it could provide explainable models, which is beneficial for critical real-world applications [7].

- **The Cold Start of Digital Twin with Prior Knowledge**

  Normally, the system needs plenty of data and time to train and tune its parameters in the digital twin. Therefore, in the code start stage of digital twin, i.e. without much data, it is usually takes plenty of efforts to tune the performance to match the real physical part. As Bayesian model, it could achieve satisfactory performance with limited training data, and a faithful prior to reflect the distribution of the weights or data [10]. The prior knowledge represented as knowledge graph could be a good choice for the cold start in digital twin system.

- **Knowledge graph incorporation to facilitate the big data processing**

  In a large-scaled digital twin, storing all sensor data is very expensive or infeasible, making it essential to efficiently acquire useful knowledge from different local sources. In this case, the research on how to represent the streaming data as knowledge, and how to utilize these knowledge in the traditional model is necessary to reduce the latency of processing to meet the requirement of digital twins. One possible way is to embed the knowledge graph as numerical vectors that acts as one input in the traditional learning model.

References

[1] Pascal Hirmer, Uwe Breitenb, and Ana Cristina. "Automating the Provisioning and Configuration of Devices in the Internet of Things." Complex Systems Informatics and Modeling Quarterly, 9:28–43, 2016.

[2] R. M. Rangan, S. M. Rohde, R. Chadha, and P. Bliznakov. "Streamlining product lifecycle processes: a survey of product lifecycle management implementations, directions, and challenges". Journal of Computing and Information Science in Engineering 5, 3 (2005), 227–237.

[3] Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management 35.2 (2015): 137-144.

[4] E. A. Feigenbaum, "Knowledge engineering: The applied side of artificial intelligence," Annals of the New York Academy of Sciences, vol. 426, pp. 91-107, Nov. 1984.

[5] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data, " IEEE Transactions Knowledge Data Engineering, vol. 26, no. 1, pp. 97-107, Jan. 2014.

[6] L. Y. Zhang, "Research on cloud data simplification," Journal of Computer Aided Design & Computer Graphics, vol. 13, no. 11, pp. 1019-1023, 2001.

[7] H. Chen, P. Tino, A. Rodan, X. Yao. "Learning in the Model Space for Cognitive Fault Diagnosis." IEEE Transactions on Neural Networks and Learning Systems, 25(1):124-136, 2014.

[8] Lu, Yuqian, et al. "Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues." Robotics and Computer-Integrated Manufacturing 61 (2020): 101837.

[9] H. Chen, P. Tino, X. Yao. "Probabilistic Classification Vector Machines." IEEE Transactions on Neural Networks, 20(6): 901-914, 2009.

[10] H. Chen, P. Tino, X. Yao. "Predictive Ensemble Pruning by Expectation Propagation." IEEE Transactions Knowledge Data Engineering, 21(7): 999-1013, 2009.

Huanhuan Chen received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2004, and the Ph.D. degree in computer science from the University of Birmingham, Birmingham, U.K., in 2008. He is currently a Full Professor with the School of Computer Science and Technology, USTC. His current research interests include neural networks, Bayesian inference, and evolutionary computation. Dr. Chen was a recipient of the 2015 International Neural Network Society Young Investigator Award, the 2012 IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award, the IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award (bestowed in 2011 and only one article in 2009), and the 2009 British Computer Society Distinguished Dissertations Award. He serves as an Associate Editor for the IEEE RANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE.