Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Internet-of-Things Enabled Manufacturing: Challenges to Machine Learning and Deep Learning

Q. Peter He*, Jin Wang**

**Department of Chemical Engineering, Auburn University, Auburn, AL 36849 USA (Tel: 334-844-7602; e-mail: qhe@auburn.edu).*
*** Department of Chemical Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: wang@auburn.edu)*

Abstract: Although many general frameworks have been proposed for Internet-of-Things (IoT) enabled systems and their potential in industrial applications, there is limited study on the properties of these industrial IoT devices. In particular, there is little study on the characteristics of the data generated from these IoT devices and what challenges they present to the machine learning and deep learning modeling techniques in industrial applications. In this work, we present our study of the data characteristics of two IoT-enabled manufacturing testbeds. One is the application of IoT vibration sensors for monitoring a pump-flow process and the other is the application of short-range IoT Wi-Fi for estimating moisture content of woodchips in a chemical pulping process. We demonstrate that the traditional data cleaning or preprocessing approaches are not good options for dealing with the veracity of IoT data. We also demonstrate that integrating domain knowledge and human learning with machine learning is the key to successful applications.

Keywords: Internet-of-Things, smart manufacturing, big data, machine learning, deep learning.

1. INTRODUCTION

The emergence of the industrial Internet-of-Things (IoT) devices and ever advancing computation and communication technologies have fueled a new industrial revolution. IoT devices are sensors, actuators and computers with wireless networks. IoT devices are typically small and easy to embed. Although the use of IoT sensors has been increasing exponentially in industries such as retail and services, their use in manufacturing has been limited. Nevertheless, there have been high expectations that IoT will be a key enabler for making manufacturing systems smarter, safer and more efficient. Because of their small size and cheap price, IoT devices offer the opportunity to instrument systems in massive numbers. With the huge amount of data and the programmability of IoT devices, comes the opportunity to shape the data received, to address local redundancy of information, and to improve both the accuracy and precision of measurements locally and across a distributed parameter system such as a reactor.

However, although many general frameworks have been proposed for IoT enabled systems and their potential in industrial applications (Bi et al., 2014; Jeschke et al., 2017; Löffler and Tschiesner, 2013; Zhong and Ge, 2018), there is limited study on the properties, such as capabilities and limitations, of these industrial IoT devices. In particular, there is little to no study on the characteristics of the data generated from these IoT devices when installed on manufacturing systems, and what challenges the collected IoT process data will present to the conventional machine learning approaches (particularly deep learning) in industrial applications such as process modeling and monitoring.

Simulation is a powerful tool, but the fidelity of the simulated system is limited by the understanding of the system. In this work, with industrial IoT still in its infancy, there is not sufficient understanding on the property, capacity and performance of IoT sensors to enable accurate simulation. In addition, most simulations only consider white noise in measurements, while other data imperfections such as outliers and missing values are seldom considered. It is well known that deep neural networks are vulnerable to adversarial samples (Papernot et al., 2016). In the case of cybersecure, the adversarial samples could be inputs crafted by adversaries with the intent of causing deep neural networks to misclassify. For example, one study shows that by perturbing only one pixel with differential evolution (DE), an image can be misclassified (one-pixel attack) (Su et al., 2019). Many other studies have shown similar results. These were the cases of manipulating testing samples to fool deep neural networks. But there is not much study on how the data imperfections (e.g., noises, outliers and missing values) in both training and testing samples would affect the performance of deep neural networks. This is a critical gap that must be filled for applying machine learning and deep learning to IoT enabled manufacturing, as data collected from industrial IoT devices will inevitably be corrupted with these imperfections. This is the major motivation for us to build smart manufacturing testbeds equipped with various IoT devices. Using the real data from these testbeds, we can investigate the challenges of the IoT process data to machine learning and deep learning, and develop strategies to address them.

2. RESULTS

In this work, we present our study of the data characteristics of two different IoT devices in two IoT-enabled manufacturing testbeds. One is the application of IoT vibration sensors for monitoring a pump-flow process (Shah et al., 2019, 2017) and the other is the application of short-range IoT Wi-Fi for estimating moisture content of woodchips in a chemical pulping process (Suthar et al., 2019).

In the application of IoT vibration sensors for monitoring a pump-flow process (Shah et al., 2019, 2017), the motor revolutions per minute (RPM) and water flow rate gallon per minute (GPM) are the key process measurements. The goal is to predict these two key process variables through IoT vibration sensor (i.e., accelerometer) measurements. Although the testbed is simple, the principles developed based on it can be readily generalized to more complex fluid flow systems.

In the application of short-range IoT Wi-Fi for estimating moisture content of woodchips in a chemical pulping process, we extract and use channel state information (CSI) from Wi-Fi signals that passing through woodchips to predict their moisture content (Suthar et al., 2019).

Our study shows that the data generated from IoT devices are truly messy big data that can be described by the 4V characteristics, namely volume, velocity, variety and veracity. We further demonstrate that the most prominent challenge that IoT data present to machine learning, especially deep learning, is the data veracity or the messiness of the data, which include a variety of components such as significant noises, irregular sampling intervals, missing values and segments.

We demonstrate that the traditional data cleaning or preprocessing approaches are not good options for dealing with the veracity of IoT data. In contrast, robust machine learning methods that do not require data cleaning or preprocessing have significant advantages. Finally, we demonstrate that in both IoT enabled applications, rote application of machine learning methods, including deep learning methods, result in underperforming models that lead to incomplete or misleading conclusions. Specifically, for the pump-flow process, direct application of long short-term memory (LSTM) recurrent neural network (RNN) to the raw vibration data results in extremely poor predictions of motor speeds and water flow rates. For the short-range IoT Wi-Fi in classifying moisture content of woodchips, direction application of various classification methods to the raw CSI data leads to significant misclassifications.

In contrast, we demonstrate that integrating human learning with machine learning is the key to successful applications. In particular, we demonstrate that robust feature engineering guided by human learning (e.g., data exploration/visualization, domain knowledge integration, and fundamental relation extraction) plays a key role for improving machine learning performance. Specifically, for the pump-flow process, with robust feature engineering, we were able to develop two simple linear machine learning models to very accurately predict both motor speed and water flow rate. For the woodchips moisture content classification application, statistics pattern analysis (SPA), a framework that we developed for robust process monitoring (He and Wang, 2011; Wang and He, 2010), was utilized to develop robust features that resulted in 100% classification accuracy.

3. CONCLUSIONS

Through two IoT-enabled manufacturing testbeds, we demonstrate that data generated from IoT devices are truly messy big data that can be described by the 4V characteristics: volume, velocity, variety and veracity. Specifically, the data veracity, particularly unequal sampling rate and missing signal segments, has the most significant impact on machine learning and deep learning. We show that the traditional data cleaning or pre-processing is not the best option for addressing IoT data veracity. Due to the very fast dynamics of these signals such as vibration or WiFi transmission, traditional data pre-processing, such as interpolation or binning, can easily cause distortion and/or loss of information contained in the signal. In comparison, we show that a preferred solution is to combine robust machine learning with robust feature engineering to directly address the data veracity without data cleaning or pre-processing. In addition, we demonstrate that complex deep learning algorithms do not necessarily perform better than simple machine learning algorithms. In particular, we show that rote application of deep learning to messy raw IoT data can result in poor models that lead to misleading results. In comparison, careful feature engineering based on domain knowledge and human learning can not only significantly improve model prediction performance but also lead to significantly simpler model that is interpretable and robust.

REFERENCES

Bi, Z., Da Xu, L., Wang, C., 2014. Internet of things for enterprise systems of modern manufacturing. IEEE Trans. Ind. informatics 10, 1537–1546.

He, Q.P., Wang, J., 2011. Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. AIChE J. 57, 107–121. https://doi.org/10.1002/aic.12247

Jeschke, S., Brecher, C., Meisen, T., Özdemir, D., Eschert, T., 2017. Industrial internet of things and cyber manufacturing systems, in: Industrial Internet of Things. Springer, pp. 3–19.

Löffler, M., Tschiesner, A., 2013. The Internet of Things and the future of manufacturing. McKinsey Co. 4.

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z.B., Swami, A., 2016. The limitations of deep learning in adversarial settings, in: 2016 IEEE European Symposium on Security and Privacy (EuroS&P). IEEE, pp. 372–387.

Shah, D., Hancock, A., Skjellum, A., Wang, J., He, Q.P., 2017. Challenges and opportunities for IoT-enabled cybermanufacturing: what we learned from an IoT-enabled manufacturing technology testbed, in: Proceedings of Foundations of Computer Aided Process Operations / Chemical Process Control. p. 66.

Shah, D., Wang, J., He, Q.P., 2019. An Internet-of-things Enabled Smart Manufacturing Testbed, in: IFAC-PapersOnLine. Florianopolis, Brazil, pp. 562–567. https://doi.org/10.1016/j.ifacol.2019.06.122

Su, J., Vargas, D.V., Sakurai, K., 2019. One pixel attack for fooling deep neural networks. IEEE Trans. Evol. Comput.

Suthar, K., Wang, J., Jiang, Z., He, Q.P., 2019. Using Channel State Information for Estimating Moisture Content in Woodchips via 5 GHz Wi-Fi, in: American Control Conference (ACC), 2020. p. Submitted.

Wang, J., He, Q.P., 2010. Multivariate Statistical Process Monitoring Based on Statistics Pattern Analysis. Ind. Eng. Chem. Res. 49, 7858–7869. https://doi.org/10.1021/ie901911p

Zhong, R.Y., Ge, W., 2018. Internet of things enabled manufacturing: A review. Int. J. Agil. Syst. Manag. 11, 126–154.