EDITORIAL

Leveraging Big Data and Machine Learning for Digital Transformation

Jingwei Huang[*]
Department of Engineering Management and Systems Engineering, Old Dominion University
Norfolk, VA 23529, USA

The fast diffusion of new technologies such as the Internet of Things (IoT), Cloud Computing, Wireless Communication, Mobile Computing, Big Data, AI (including Machine Learning), Cyber-Physical Systems, Blockchain, and others, has been leading to a pervasive digital transformation (Foster, 2020; Huang, 2017). In this transformation, many traditional artifacts and business processes are digitalized. Examples include digital medical records, digital media, digital currencies, digital twin, digital manufacturing, and digital engineering (US DoD, 2018). Beyond the digital forms of artifacts and processes and the associated big data phenomenon, the digitalization enables fast dissemination of digital information and much-enhanced knowledge sharing, enables virtualization of various services, enables innovations by leveraging digital technologies, and fosters the creation of new "data products" and "digital knowledge products". Digital transformation is reshaping the landscape of systems design (Huang et al., 2020). Digital transformation and the associated technologies are efficiently and effectively helping the world to cope with the challenging Covid-19 pandemic. On the other hand, they have also been impacting many aspects of the fast-transforming human society and lead to many issues to deal, to think, and to explore.

In today's rapid-changing and highly competitive environment with an unprecedented richness of information, it is an essential capability to turn big data into knowledge, in order to support effective and efficient decision making and operations processes, no matter in a field of scientific research, engineering, business, healthcare, public services, or others. In the transformation of data into knowledge, machine learning and data analytics play a central role, where efficient big data platforms are fundamental for handling the big data. This issue of JIDPS presents four papers, reflecting research topics on the aspects of deep neural network learning, data analytics, resource management of big data platforms, and automation of multimedia data publishing.

In the last decade, thanks to the availability of big data and the increase of computing power, including GPUs (Graphics Processing Units) and cloud computing enabled big data platforms, machine learning has made significant progresses and has been broadly applied in many fields. One of the representative machine learning technologies is deep learning with convolutional neural networks (CNNs) (LeCun, Bengio, & Hinton, 2015). CNN is first proposed by LeCun and others in (Lecun, Bottou, Bengio, & Haffner, 1998) and later becomes populated by the "AlexNet" (Krizhevsky, Sutskever, & Hinton, 2012). VGGNet (Simonyan & Zisserman, 2014) is a family of successful and influential works that demonstrate the power

---

[*] Corresponding author. Email: j2huang@odu.edu.

of deep CNNs, conducted by Visual Geometry Group (VGG) at Oxford University. Among them, VGG16 (Simonyan & Zisserman, 2014) is an award-winning deep CNN model instance. It was trained with over 1.3 million images in 1000 categories, by using four NVIDIA Titan Black GPUs and taking 2~3 weeks. Deep learning is computing-intensive and associated with high cost with respect to computing resources, time, and labour on both experiments and labelling samples. Therefore, it is important to find ways to reuse the knowledge learned in previous works. "Transfer learning" (Pan & Yang, 2009; Weiss, Khoshgoftaar, & Wang, 2016) is a vehicle towards that end. Basically, CNNs are characterized by the strong capability of feature identification from signals by using filters and can be used for various classification problems. Since an object of interest may have features in different sizes, interestingly, in recent years, multiscale filer CNNs (i.e., CNNs with multiple filters of different sizes) (Zhao & Jia, 2016) have been proposed to address this type of problem. Article "Deep Learning for the Automatic Diagnosis and Detection of Multiple Retinal Abnormalities", by Michelle K.S. Gian, Valliappan Raman, and Patrick H.H. Then, demonstrates the application of those up-to-date technologies from CNN deep learning for retinal abnormalty detection, classification, and segmentation. They reuse the "general purpose" model VGG16 as an "off-the-shelf" feature extractor (encoder) in their deep CNN architecture, and re-train the CNN with a limited number of available labelled retinal images to fine-tune the pre-trained model towards the specific purpose of retinal disease classification. This application also reflects an interesting perspective regarding how digitalized artifacts can be used/reused to transfer data into knowledge and the potential of data products and digital knowledge products in a digitalized world.

Big data offer much richer information for studying complex systems, including social systems, thus enabling the observation of the object of interests from different views and leveraging additional data for the exploration of new approaches different from the traditional ones. Article "Time Reversed Delay Differential Equation Based Modelling of Journal Influence in An Emerging Area", by Archana Mathur, Snehanshu Saha , Poulami Sarkar, Saibal Kar, and Suryoday Basak, presents the authors' study on journal ranking with respect to the growth of new journals in emerging fields by using delay differential equations and incorporating factors of editors' reputation and publishers' goodwill.

Cloud Computing is today's mainstream of big data platforms to facilitate big data gathering, storage, processing, sharing, publishing, and machine learning. Therefore, it is critical to maintain the trustworthiness of cloud computing with respect to performance, security, privacy, and fairness. Among those, reliable performance is essential. A cloud data centre may have dozens of thousands of machines in several clusters, and each cluster can have above ten thousand machines. As examples, a Borg cluster at Google has 12 thousands of nodes (Tirmazi et al., 2020) and a Mesos cluster at Twitter has over ten thousands of nodes (Hashemi, 2017). Given such a large number of machines, it is almost certain that many faults pop up frequently, thus how to effectively and efficiently operate in a fault-prone operating environment poses a significant challenge. Article "Fault Tree Analysis based Virtual Machine Migration for Fault-Tolerant Cloud Data Center", by Getzi Jeba Leelipushpam, Immanuel Johnraja Jebadurai, and Jebaveerasingh Jebadurai, presents the authors' research on virtual machine migration strategy based on estimated job makespan and the mean time to failure with the support of a fault tree analysis.

In the era of big data, it is a common task to integrate information from different sources and to dynamically compose and publish various forms of data. Therefore, it is important to have the capability of the automagical composition of hypermedia documents, to make them presentable in a coherent manner. As examples, there is no overlap between pictures and texts, and no more than one video or audio files are playing. Article "Synchronization Management in Hypermedia Documents", by Naima Beladam and Abdelghani Ghomari, presents their research on a formal approach to hypermedia document composition and verification and the development of an experimental system, called "Hypermedia Document Builder (HDB)". Their work is based on Separation Logic, an extension to Hoare Logic, which allows formal specification and verification of the temporal and spatial constraints between multimedia objects in a hypermedia document.

By this chance, the journal would like to specially thank Dr. J. Dinesh Peter for his contribution in recommendation and organizing the review of papers "Time Reversed Delay Differential Equation Based

Modelling of Journal Influence in An Emerging Area" and "Fault Tree Analysis based Virtual Machine Migration for Fault-Tolerant Cloud Data Center".

No doubt, with the deeper and broader digitalization, many innovative applications and new technologies on big data and machine learning will emerge. JIDPS will continue to publish articles on this theme and relevant emerging research topics.

References

Foster, M. (2020). The 360-Degree Approach to Digital Reinvention. *MIT Sloan Management Review*. Retrieved from https://sloanreview.mit.edu/audio/the-360-degree-approach-to-digital-reinvention/

Hashemi, M. (2017). The Infrastructure Behind Twitter: Scale. Retrieved August 3, 2020, from Twitter Blog website: https://blog.twitter.com/engineering/en_us/topics/infrastructure/2017/the-infrastructure-behind-twitter-scale.html

Huang, J. (2017). Building Intelligence in Digital Transformation. *Transactions of the SDPS: Journal of Integrated Design and Process Science*, (Preprint), 1–4.

Huang, J., Gheorghe, A., Handley, H., Pazos, P., Pinto, A., Kovacic, S., … Daniels, C. (2020). Towards Digital Engineering -- The Advent of Digital Systems Engineering. *International Journal of System of Systems Engineering*. Retrieved from http://arxiv.org/abs/2002.11672

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Tirmazi, M., Barker, A., Deng, N., Haque, M. E., Qin, Z. G., Hand, S., … Wilkes, J. (2020). Borg: the next generation. *Proceedings of the Fifteenth European Conference on Computer Systems*, 1–14.

US DoD. (2018, June). *Digital Engineering Strategy, Office of the Deputy Assistant Secretary of Dense for Systems Engineering*. Retrieved from https://fas.org/man/eprint/digeng-2018.pdf

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9.

Zhao, L., & Jia, K. (2016). Multiscale CNNs for brain tumor segmentation and diagnosis. *Computational and Mathematical Methods in Medicine*, *2016*.