

# Data Analytics for IoT Applications

Yuh-Jye Lee

Department of Computer Science and Information Engineering  
National Taiwan University of Science and Technology

Taipei, Taiwan

e-mail: [yuh-jye@csie.ntust.edu.tw](mailto:yuh-jye@csie.ntust.edu.tw)

***Abstract***—IoT applications are expected to have huge impact to us in a very near future. People imagine Intelligent Transportation Systems, Intelligent Care Systems, smart buildings, or even smart cities may become reality to facilitate our daily lives from the moment we wake up from beds to the time we go to sleep. To efficiently and effectively manage various IoT applications, we should carefully check the data that are generated by IoT devices to understand the environment states and monitor any particular anomalous conditions if there is any. Data driven approach becomes a key corner stone for the success of most IoT devices. Compared to traditional data analytics research, data analysis on IoT applications seems to be more challenging simply due to the huge amount of data that can be easily generated by IoT devices, diverse data format that could be generated and rapid data that could be generated in a small period. Moreover, data analysis for IoT data usually is worked upon different assumptions as we usually have limited computing resources, limited power to be used and limited bandwidth to send data in IoT applications. That makes data analytics hard to operate especially when the task must be done online or in (close to) real time.

In this talk, we survey a list of data analytics methodology that is appropriate to be used for IoT applications, following a discussion on the difference between traditional data analytics and IoT data analytics. Anomaly detection can also be considered as an important step for IoT data analysis. For IoT applications with limited resources such as power or bandwidth, we can choose the analytics model that is simple but good enough as an approximate version of the exact model for knowledge extraction given IoT data. We can also use distributed modeling techniques for IoT data analysis because the IoT applications may be deployed in a distributed manner. On distributed models, we often consider spatial or temporal relationship or both between different IoT data to enhance the understanding of the IoT data. Some applications will be addressed in this talk too.

*About the Speaker:* Dr. Yuh-Jye Lee received the PhD degree in Computer Science from the University of Wisconsin-Madison in 2001. He is currently a Professor of Department of Computer Science and Information Engineering at National Taiwan University of Science and Technology. He also serves as a principal investigator at the Intel-NTU Connected Context Computing Center. His research is primarily rooted in optimization theory and spans a range of areas including network and information security, machine learning, big data, data mining, numerical optimization and operations research. During the last decade, Dr. Lee has developed many learning algorithms in supervised learning, semi-supervised learning and unsupervised learning as well as linear/nonlinear dimension reduction. His recent research is applying machine learning to information security problems such as network intrusion detection, anomaly detection, malicious URLs detection and legitimate user identification. Currently, he focuses on online learning algorithms for dealing with large scale datasets, stream data mining and behavior based anomaly detection for the needs of big data, Internet of Things data analytics and machine to machine communication security problems