SysML: On System and Algorithm Co-design for Practical Machine Learning

Eric Xing
Carnegie Mellon University and Petuum
Pittsburgh, PA
epxing@cs.cmu.edu

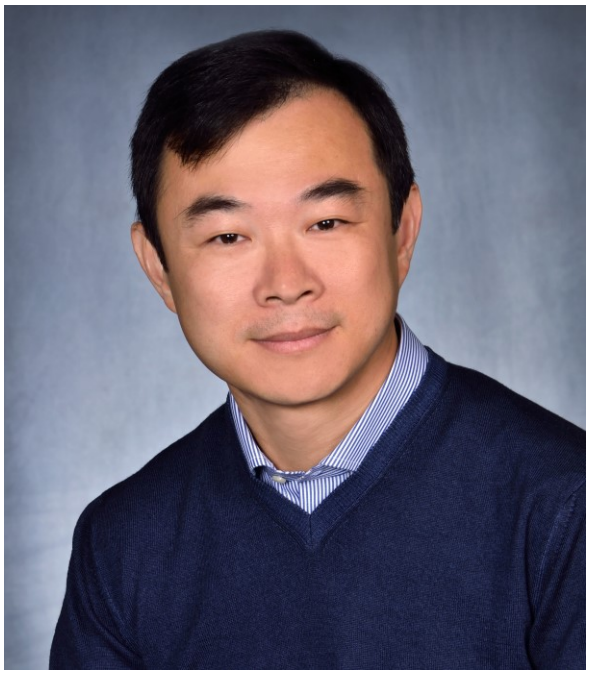
ABSTRACT

The rise of Big Data and AI computing has led to new demands for Machine Learning systems to learn complex models with millions to billions of parameters that promise adequate capacity to digest massive datasets and offer powerful and real-time predictive analytics thereupon. In this talk, I discuss a recent trend toward building new distributed frameworks for AI at massive scale known as "system and ML algorithm co-design", or SysML -- system designs are tailored to the unique properties of ML algorithms, and algorithms are re-designed to better fit into the system architecture. I show how one can explore the underlying statistical and algorithmic characteristics unique to ML programs but not typical in traditional computer programs in designing the system architecture to achieve significant, universal, and theoretically sound power-up of ML program across the board. I also present a briefly introduction of the Petuum system based on such interdisciplinary innovations, which intends to dramatically improve adoption of AI solutions by lowering the barrier of entry to AI technologies via Automatic Machine Learning through Petuum. I show how, through automatable, product-grade, hardware-agnostic, standardized building blocks that can be assembled and customized, AI users can liberate themselves from the demanding experience of algorithm programming and system tuning, and easily experiment with different AI methods, parameters, and speed/resource trade-offs by themselves or automatically.

To put this in a broader context, recent discussions about AI in both research community, and the general public have been championing a novelistic view of AI, that AI can mimic, surpass, threaten, or even destroy mankind. And such discussions are fueled by mainly recent advances in deep learning experimentations and applications, which are however often plagued by its craftiness, un-interpretability, and poor generalizability. I will discuss a different view of AI as a rigorous engineering discipline and as a commodity, where standardization, modularity, repeatability, reusability, and transparency are commonly expected, just as in civil engineering where builders apply principles and techniques from all sciences to build reliable constructions. I will discuss how such a view sets different focus, approach, metric, and expectation for AI research and engineering, which we practiced in our SysML work.

BIOGRAPHY

Eric Xing is a Professor of Computer Science at Carnegie Mellon University, and Founder and CEO of the machine learning platform startup Petuum Inc. He completed his undergraduate study at Tsinghua University, and holds a PhD in Molecular Biology and Biochemistry from the State University of New Jersey, and a PhD in Computer Science from the University of California, Berkeley. His main research interests are the development of machine learning and statistical methodology, and large-scale computational system and architectures, for solving problems involving automated learning, reasoning, and decision-making in high-dimensional, multimodal, and dynamic possible worlds in artificial, biological, and social systems. Prof Xing is a board member of the International Machine Learning Society; in 2014, he served as the Program Chair of the International Conference of Machine Learning (ICML), and in 2019, he will serve as the General Chair of ICML. He is the Associate Department Head of the Machine Learning Department at Carnegie Mellon University, and a Fellow of the Association of Advancement of Artificial Intelligence.



KDD 2018, August 19-23, 2018, London, United Kingdom.

ACM ISBN 978-1-4503-5552-0/18/08.
DOI: https://doi.org/10.1145/3219819.3219934