ParLearning 2017 Keynote 1

Why Tables and Graphs for Knowledge Discovery Systems

John Feo

Northwest Institute for Advanced Computing, Pacific Northwest National Laboratory and University of Washington, USA

Abstract: The availability of data is changing the way science, business, and law enforcement operate. Economic competitiveness and national security depend increasingly on the insightful analysis of large data sets. The breadth of analytic processes is forcing knowledge discovery platforms to supplement traditional table-based methods with graph methods that provide better support for sparse data and dynamic relationships among typed entities. While storing data in only tables makes it difficult to discover complex patterns of activities in time and space, tables are the most efficient data structure for storing dense node and edge attributes, and executing simple select and join operations. Consequently, knowledge discovery systems must support both in a natural way without preference. In this talk, I will describe the hybrid data model, SHAD, that we are developing to support both graphs and table. I will present several high-performance, scalable, analytic platforms developed by PNNL for graph analytics, machine learning, and knowledge discovery. I will conclude with an overview of HAGGLE, our proposed graph analytic platform for new architectures.

Speaker: Dr. John Feo is Director of the Northwest Institute for Advanced Computing, a joint institute established by Pacific Northwest National Laboratory and University of Washington. Previously, he managed a large DOD research project in graph algorithms, search, parallel computing, and multithreaded architectures. Dr. Feo received his Ph.D. in Computer Science from The University of Texas at Austin. He began his career at Lawrence Livermore National Laboratory where he managed the Computer Science Group and was the principal investigator of the Sisal Language Project. Dr. Feo then joined Tera Computer Company (now Cray Inc.) where he was a principal engineer and product manager for the MTA-1 and MTA-2, the first two generations of Cray’s multithreaded architecture. He has taken short sabbaticals to work at Sun Microsystem, Microsoft, and Context Relevant. Dr. Feo has held academic positions at UC Davis and Washington State University.
ParLearning 2017 Keynote 2

Matrix Factorization on GPUs: A Tale of Two Algorithms

Wei Tan

IBM T. J. Watson Research Center, NY, USA

Abstract: Matrix factorization (MF) is an approach to derive latent features from observations. It is at the heart of many algorithms, e.g., collaborative filtering, word embedding and link prediction. Alternating least Square (ALS) and stochastic gradient descent (SGD) are two popular methods in solving MF. SGD converges fast, while ALS is easy to parallelize and able to deal with non-sparse ratings. In this talk, I will introduce cuMF, a CUDA-based matrix factorization library that accelerates both ALS and SGD to solve very large-scale MF. cuMF uses a set of techniques to maximize the performance on single and multiple GPUs. These techniques include smart access of sparse data leveraging memory hierarchy, using data parallelism with model parallelism, approximate algorithms and storage. With only a single machine with up to four Nvidia GPU cards, cuMF can be 10 times as fast, and 100 times as cost-efficient, compared with the state-of-art distributed CPU solutions. In this talk I will also share lessons learned in accelerating compute- and memory-intensive kernels on GPUs.

Speaker: Dr. Wei Tan is a Research Staff Member at IBM T. J. Watson Research Center. His research interest includes big data, distributed systems, NoSQL and services computing. Currently he works on accelerating machine learning algorithms using scale-up (e.g., GPU) and scale-out (e.g., Spark) approaches. His work has been incorporated into the IBM patent portfolio and software products such as Spark, BigInsights and Cognos. He received the IEEE Peter Chen Big Data Young Researcher Award (2016), Best Paper Award at ACM/IEEE ccGrid 2015, IBM Outstanding Technical Achievement Award (2017, 2016 and 2014), Best Student Paper Award at IEEE ICWS 2014, Best Paper Award at IEEE SCC 2011, Pacesetter Award from Argonne National Laboratory (2010), and caBIG Teamwork Award from the National Institute of Health (2008).