**Accelerating Machine Learning with Model-Centric Approach on Emerging Architectures**

Judy Qiu, Indiana University

Learning from massive datasets has applications in various scientific and engineering disciplines, with a variety of data such as text, images, and that from high-throughput biology. The scale of these datasets, however, is often prohibitive on single node computer. An obvious approach is to parallelize machine learning algorithms. There is much work right now on using HPC machines (e.g. many-core and GPU) for Big Data applications such as deep learning and machine learning. Increasingly we see novel software architectures that are important because they make the optimization problem easier, like deep learning frameworks (e.g. Caffe, Torch, and Tensorflow) and machine learning frameworks (e.g. Hadoop, Harp, Spark, Flink, GraphLab, Parameter Server, and Petuum). We propose a systematic approach to the parallelization of machine learning algorithms. A central idea is to distinguish the (input) data and model (parameter) components of the algorithm and design a runtime and programming paradigm that supports this. Key to our model-centric approach is a categorization of parallel machine learning algorithms into four types of computation models. We have shown that previous standalone enhanced versions of MapReduce can be replaced by Harp (a Hadoop plug-in) that offer both data abstractions useful for high-performance iterative computation and MPI-quality communication. We select a subset of machine learning algorithms, Latent Dirichlet Allocation using Collapsed Gibbs Sampling and Matrix Factorization using Stochastic Gradient Descent and implement them with optimized performance using Hadoop/Harp as a distributed framework to invoke Intel’s Data Analytics Acceleration Library (DAAL) and describe experimental results on Intel’s Haswell and Knights Landing architectures.

**Short Bio**

Dr. Judy Qiu is an associate professor of Intelligent Systems Engineering in the School of Informatics and Computing at Indiana University. Her research interests are parallel and distributed systems, cloud computing, and high-performance computing. She leads the SALSA project, encompassing data-intensive computing at the intersection of Cloud and multicore technologies, and offers an online course CloudMOOC as part of the Data Science Program of the School of Informatics and Computing. Her research has been funded by NSF, NIH, Intel, Microsoft, Google, and Indiana University. Judy Qiu leads the Intel Parallel Computing Center (IPCC) site at IU. She is the recipient of a NSF CAREER Award in 2012, Indiana University Trustees Award For Teaching Excellence in 2013-2014, and Indiana University Outstanding Junior Faculty Award 2015.