

## Abstract

# Performance Modeling of Task Parallel Programs

BYAMBAJAV NAMSRAIJAV<sup>1,a)</sup> KENJIRO TAURA<sup>1,b)</sup>

Presented: January 10, 2017

To estimate the execution time of an application for given parameters on a given hardware, people often use analytical models if they are familiar with the application logic. But such approaches become increasingly difficult as the application becomes complicated and the number of input parameters increase. Besides for task parallel applications, the performance is highly dependent on the characteristics of the underlying dynamic task scheduler runtime. Therefore, it is extremely challenging to formulate analytical models for task parallel applications as the performance depends on not only the application logic but also the underlying hardware properties and the task parallel runtime's performance. Machine learning techniques can be used to build a performance model when formulating reliable analytical model is unfeasible. First, we run the target application multiple times to learn the execution time for differing input parameters and worker counts. Then a machine learning model is built using that training data. However, past such models, mainly developed for well load-balanced applications, perform poorly when applied to task parallel applications where there is much load-imbalance. Also, the accuracy of those models significantly decreases when the number of workers used in the prediction target execution becomes bigger than the number of maximum workers used during the training. We investigate whether applying modern machine learning techniques can address these challenges. In this presentation, we build performance models for task parallel programs using Lars-Lasso and deep neural network regression. We also exploit some additional information which we can gather during the training, such as the hardware performance counter values, the number of tasks created, and the task-stealing overhead, to guide these models. We evaluate the proposed models for BOTS and PARSEC benchmark applications running on multiple task parallel runtimes.

---

<sup>1</sup> Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan

a) byambajav@idos.ic.i.u-tokyo.ac.jp

b) tau@idos.ic.i.u-tokyo.ac.jp