Parallel Hoeffding Tree on GPU

Chongyang Xu
Beihang University
xuchongyang1995@buaa.edu.cn

Abstract

Decision trees are widely used machine learning methods in practice. Due to the increasing data volume and requirement of learning quickly from streaming data, many parallelizing methods are proposed to meet the increasing computation requirement. GPU devices have high parallelism; however, much research explores parallelizing on CPUs. Moreover, the only work on GPU occupies much memory, thus limiting the depth of decision trees.

The challenges of parallelizing online decision trees on GPUs include 1) lacking efficient storage structures of decision trees on GPU, 2) low parallelism in computation.

This project has carried out the following three aspects on GPUs. (1) Parallelizing online decision tree-Hoeffding tree with a compact structure and exploiting multi-level parallelism. (2) Providing an ensemble of Hoeffding trees. (3) Implementing a prototype system for online decision tree learning.

Experiment shows that the GPU accelerated Hoeffding tree has comparable statistic performance to CPU version and can reach 3.61~10.87 times the speed of CPU version, and the depth of tree can reach 194. Online bagging decision tree and online random forest has comparable statistic performance and can reach 1.34~4.48 and 0.89~4.29 times the speed of CPU version, respectively.