

## Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: RNDr. Martin Kruliš, Ph.D.

Jméno a příjmení autora práce: Bc. František Haas

Název práce: Parallel Processing of Huge Astronomical Data

---

Práce se zabývá paralelizací metody náhodných lesů (Random Forests) na GPU. Náhodný les je metoda strojového učení používaná pro klasifikaci a regresi. Autor implementoval tento algoritmus pomocí platforem OpenCL a CUDA a empiricky ověřil výkon navržených prototypů.

Po obecném úvodu do problematiky popisuje text práce algoritmus náhodných lesů a principy programování GPU pomocí OpenCL a CUDA. Popis samotného algoritmu je podle mého názoru příliš stručný a řada použitých značení není vysvětlena. V některých situacích je navíc použito různé označení či názvosloví pro totéž (např. variables vs. attributes).

Čtvrtá kapitola popisuje základní implementaci a její testy v OpenCL, pátá kapitola se věnuje první verzi napsané pomocí CUDA a šestá kapitola popisuje vylepšenou verzi CUDA implementace, která odstraňuje problémy s režii častého volání CUDA API funkcí. OpenCL a CUDA verze na sebe příliš logicky nenavazují a používají jiný přístup k implementaci. Není tedy jasné, jaký je přínos ohledně jejich srovnání. Dále v popisu chybí množství detailů (konkrétní počty spuštěných vláken, uspořádání vláken do skupin, přesné uspořádání dat v paměti, ...), které jsou u GPU implementací poměrně zásadní, protože značně ovlivňují celkový výkon. V neposlední řadě bych poznamenal, že těmto kapitolám chybí podrobnější analýza, která by vysvětlovala, proč autor zvolil popsání postupy a proč se domnívá, že jsou to postupy nejlepší možné.

Sedmá kapitola popisuje samotnou klasifikaci astronomických dat, což byla původní motivací práce. Osmá kapitola stručně sumarizuje existující výzkumné práce na podobná témata. V závěru práce pak autor navrhuje možná vylepšení, která by podle mého názoru měla být vyřešena ještě v rámci prezentovaných implementací.

Příložené zdrojové kódy svým rozsahem a strukturou odpovídají experimentální práci. Autor na nich demonstroval, že je schopen vytvořit SW prototyp i pro složitější paralelní architekturu jako je právě GPU. Jediný zásadní problém je, že jsem nenalezl zdrojové kódy OpenCL implementace, která je prezentovaná v kapitole 4. Vzhledem k tomu, že v empirickém srovnání dopadla OpenCL implementace nejlépe, je její absence přinejmenším překvapivá. Věřím, že se však jedná o čistě formální nedostatek, a autorovi bych doporučil, aby OpenCL implementaci doložil při obhajobě.

### Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

V Praze dne: 1. 9. 2016

Podpis: