

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Bc. Tomáš Malíček
Oponent práce: Ing. Ondřej Stuchlík
Název práce: Parallel Implementation of Symbolic Regression
Obor: Znalostní inženýrství

Datum vytvoření: 6. 6. 2017

Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 5:
1. Náročnost a další komentář k zadání	1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání
Popis kritéria: Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)	
Komentář: Zadání považuji za náročnější, jelikož spojuje teorii symbolické regrese, která se opírá o evoluční algoritmy, konkrétně genetické programování, s paralelizací na výpočetním clusteru.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
2. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.	
Komentář: Teoretická část je splněna zcela. V experimentální části jsou určité nedostatky, viz níže.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
3. Rozsah písemné zprávy	1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky
Popis kritéria: Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
Komentář: Celkově je písemná zpráva dostatečná, pouze kapitola 4 (implementace) postrádá (alespoň) konceptuální popis přístupu k paralelizaci koevoluce.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	100 (A)
Popis kritéria: Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře.	
Komentář: Práce je logicky dobře členěna - teoretická část správně postupuje od vysvětlení obecného genetického algoritmu ke genetickému programování, kterého je využito v implementaci symbolické regrese. Praktická část vhodně navazuje implementací a dále popisem experimentálních výsledků.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
5. Formální úroveň práce	90 (A)
Popis kritéria: Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 14/2015, článek 3.	
Komentář: Po typografické stránce je práce v pořádku. Vytkl bych pouze několik chyb v anglickém jazyce, především v kapitolách 4 a 5.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
6. Práce se zdroji	100 (A)

Popis kritéria:

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Student pracoval s dostatečným množstvím zdrojů, práce s literaturou je na výborné úrovni a citace jsou používány správně.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

7. Hodnocení výsledků, publikační výstupy a ocenění

70 (C)

Popis kritéria:

Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

Komentář:

Výsledky v oblasti základní symbolické regrese jsou předkládány neúplně. Je uváděno zrychlení, ovšem bez uvedení počtu procesorů, které byly k dosažení tohoto zrychlení použity (tzn chybí metrika zrychlení na procesor). V části zabývající se koevolucí je nevhodně zafixován parametr Predictorů určující počet řádků datasetu využívaných pro odhad chyby, stejně tak jako parametr určující velikost populace Predictorů a Trainerů.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

Uveďte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uveďte možnosti využití výsledků ZP v praxi.

Komentář:

Teoretická část je dobře využitelná jako materiál pro popis použití genetického algoritmu pro implementaci symbolické regrese. Experimentální část nevhodně porovná symbolickou regresi bez a s využitím koevoluce, zde by bylo zapotřebí porovnat i absolutní čas běhu a dosaženou chybu.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

9. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

- 1) V závěru je uvedeno: "...the populations of predictors and trainers have to be evaluated. These populations are very small (eight predictors and ten trainers in this case)..". Proč jsou populace zvoleny takto malé? Je možné zvolit velikosti těchto populací větší, případně jaké dopady by tato změna měla (v kontextu paralelizace a dosažené chyby)?
- 2) V návaznosti na tvrzení "... the fitness predictors require an extra time for its evolution which is difficult to parallelize.", v čem je paralelizace evoluce prediktorů náročná?
- 3) Bylo by možné, případně jak, upravit nastavení parametrů koevoluce tak, aby se i v jejím případě vyplatila paralelizace?

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

10. Celkové hodnocení

80 (B)

Popis kritéria:

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nesmí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení:

Teoretická část práce je velmi dobře zpracována, dobře se čte a může posloužit jako vhodný zdroj pro získání celkového přehledu o použití genetického programování jako nástroje pro symbolickou regresi. Praktická část obsahuje určité nedostatky. Zrychlení algoritmu je uváděno bez návaznosti na počet procesorů, které byly pro dosažení zrychlení použity. Závěr o nevhodnosti paralelního vyhodnocování koevoluce považuji za nepřesný, jelikož koevoluce není nevhodná pro paralelizaci per se, nevhodná se jevila pouze ve světle zvolených koevolučních parametrů. Porovnání symbolické regrese s a bez použití koevoluce bylo zkreslující, jelikož nezahrnovalo absolutní čas běhu a dosaženou chybu.

Podpis oponenta práce: