

Posudek oponenta bakalářské práce

Student: Doležal Petr
Téma: Porovnání variant genetického programování v úloze symbolické regrese (id 24804)
Oponent: Hurta Martin, Ing., UPSY FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Jedná se o mírně obtížnější zadání vyžadující nastudování a implementaci několika variant genetického programování a následné provedení statistického vyhodnocení na úloze symbolické regrese.
- 2. Splnění požadavků zadání** **zadání splněno**
Student splnil všechny body zadání.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
Technická zpráva odpovídá obvyklému rozsahu.
- 4. Prezentací úroveň předložené práce** **65 b. (D)**
Technická zpráva se dobře čte a na pár výjimek jednotlivá témata i vhodně popisuje. Mezi výjimky patří některá tvrzení v 2. kapitole, která jsou diskutabilní či nejasná. Například, že běžně se v koevolučních algoritmech pracuje s populacemi prediktorů fitness a archivem trenérů fitness. Dalším příkladem je tvrzení, že pokud má potomek u CGP stejnou fitness jako jeho rodič, je tento potomek zvolen jako rodič příští generace. Což neplatí bez uvedení důležitého faktu, že se toto týká případu, kdy potomek i rodič mají shodně nejvyšší fitness v populaci.

Jednotlivé kapitoly na sebe vhodně navazují, struktura a rozsah některých kapitol však není vhodný. Kapitola 3, popisující symbolickou regresi řešenou touto prací, má pouze jednu stranu. Návrh nástroje, umístěný v kapitole 4, má pak pouze dvě strany, a slouží spíše jako úvod k implementaci než popis návrhu.

Implementace je v kapitole 5 popsána dobře a jasně. Tabulka 5.1 obsahující výčet všech čísel od -20 do 20 s krokem 0,2 je však nadbytečná. Obrázky 5.2, 5.4 a 5.5 prezentující podobu vstupu a výstupů nástroje formou snímků obrazovky s bílým textem na černém pozadí nejsou zrovna přívětivé pro tisk ani čtenáře. Všechny obrázky v rámci práce (včetně grafů a schémat) pak využívají rastrovou grafiku a v řadě případů zabírají více místa, než by bylo odpovídající.

Práce obsahuje velkou řadu dobře vyhodnocených experimentů. V prezentaci nových výsledků je však pokračováno i v kapitole 7, diskuse, která je tedy spíše pokračováním kapitoly o experimentech. Dojem z obou kapitol bohužel kazí opakované střídání minulého, přítomného a budoucího času. Dalším pochybením je pak neuvedení hardwarových parametrů systému využitého pro běh experimentů.
- 5. Formální úprava technické zprávy** **70 b. (C)**
Text práce je psán dobře, s občasným delším souvětím nebo méně formálním úvodem kapitoly 3. Práce obsahuje pouze malé množství překlepů či gramatických chyb.

Hlavním typografickým prohřeškem je nadměrné využívání číslovaných kapitol třetí úrovně, jak napovídá obsah zabírající přes dvě strany práce. V důsledku tohoto rozhodnutí dochází k umístění až pěti podkapitol na jednu stranu. Kapitola 4 pak využívá pouze nadpisy třetí úrovně s přeskočením druhé.

Mezi další chyby patří opakované využívání desetinné tečky místo čárky nebo opomenutí tečky na konci popisů obrázků.
- 6. Práce s literaturou** **75 b. (C)**
Student čerpal celkem z 19 odpovídajících zdrojů. Jedná se z většiny o dobře zvolené monografie či vědecké články. Zdroje jsou v práci řádně citovány.

Opomenutím je absence uvedení existujících poznatků z literatury o porovnání jednotlivých variant genetického programování.
- 7. Realizační výstup** **80 b. (B)**
Výstupem práce je nástroj v jazyce Java umožňující provedení symbolické regrese za pomoci jednoho ze čtyř typů genetického programování. Nástroj přijímá jako vstup data ve formátu .CSV s libovolným množstvím

nezávislých proměnných a jedné závislé proměnné. Nástroj zároveň umožňuje automatické provedení více běhů všech variant a získání dat o dosažené fitness či době běhu.

Zdrojové kódy jsou dobře strukturovány i správně dokumentovány. Výsledný program je pak nejen funkční ale i jednoduše použitelný.

8. Využitelnost výsledků

Výsledky experimentů potvrzují aktuální vědění a doplňují jej o porovnání stromové reprezentace genetického programování a kartézského genetického programování při využití koevoluce prediktorů fitness v obou z nich. Vytvořený nástroj je možné využít k symbolické regresi. Jedná se však především o nástroj cílený k férovému porovnání jednotlivých variant a zvolený způsob implementace limituje jeho využití pro řešení výpočetně náročnějších úloh.

9. Otázky k obhajobě

- U implementace CGP zmiňujete využití evoluční strategie (1+lambda), kterou jste ale upravil doplněním jednoho náhodně vygenerovaného jedince. Máte toto rozhodnutí podloženo experimentálními výsledky?
- Proč se podle vás ukázal rekursivní výpočet CGP rychlejší než běžný postup založený na zjištění aktivních uzlů?
- V kapitole 7 sám zmiňujete, že jednotlivé krabicové grafy zobrazují rozdílné výsledky i pro shodná nastavení parametrů. Jak by šlo tento jev omezit?

10. Souhrnné hodnocení

75 b. dobře (C)

Práce se podrobně zabývá hned čtyřmi variantami genetického programování, které byly implementovány, hojně statisticky vyhodnoceny a navzájem porovnány. Kvůli míře odvedené práce a užitečným poznatkům plynoucím z provedených experimentů tak navrhuji, i přes výše uvedené výtky k některým částem technické zprávy, hodnocení **C - dobře**.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 20. května 2022

Hurta Martin, Ing.
oponent