Posudek oponenta závěrečné práce

České vysoké učení technické v Praze **Fakulta informačních technologií**

Student: Bc. Vojtěch Hejl
Oponent práce: Ing. Jan Žegklitz
Název práce: Selection of surrogate models for evolutionary black-box optimization in noisy environment
Obor: Znalostní inženýrství

Datum vytvoření: 27. 1. 2018

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:
Práce se zabývá algoritmem CMA-ES, známým algoritmem pro black-box optimalizaci, a použitím náhradních modelů pro jeho urychlení ve smyslu počtu vyhodnocení optimalizované funkce. Práce je součástí většího projektu, který se věnuje náhranímu modelování a je založena na předchozím výzkumu v rámci tohoto projektu.

Zadání má průměrnou náročnost.

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:
Zadání, tak, jak je dáno, bylo splněno - student podropbně popisuje algoritmus CMA-ES, čímž dokazuje své porozumění tomuto algoritmu; popisuje a vyhodnocuje jeho existující varianty s náhradním modelováním na několika testovacích úlohách a navrhuje rozšíření těchto variant; implementuje tato rozšíření ve zdrojovém kódu, který pracuje jako součást většího frameworku, který vzniká v rámci projektu, s nímž se student musel seznámit; testuje svá rozšíření a vyhodnocuje jejich výsledky. Nicméně ke splnění zadání mám několik výhrad. Největší z nich je způsob, jakým jsou prezentovány a vyhodnoceny výsledky.

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:
Celkový rozsah písemné části práce je adekvátní. Nicméně poměr textu věnující se samotným přínosům práce vůči základům (zejména popisu algoritmu CMA-ES) by mohl být větší.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Věcná a logická úroveň práce	*52 (E)*

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:
Celková věcnost a logičnost práce není zlá - text plyne smysluplně a má logickou strukturu. Úvodní část přednáší téma a výzkumné otázky, střední část popisuje základy a vlastní přínos práce, závěr odpovídá na výzkumné otázky a shrnuje výsledky a přínosy. Nicméně, je zde několik problémů, některé z nich poměrně závažné:

- Ve druhé odrážce na konci Introduction je výzkumná otázka: "Which alternative of CMA-ES is better?" Když se použije druhý stupeň přídavného jména (tzn. "better", tj. "lepší"), musí se říct, s čím se má srovnávat (lepší než co?) a na základě jakých kritérií (lepší v čem?). Toto zde není uvedeno a způsobouje to problémy při odpovídání na tuto otázku v Conclusion.
- V sekci 2.3, která se zabývá volbou parametrů na základě testovacích běhů, je několikrát uvedeno, že hodnota X byla nejlepší, ale nebyla vybrána, aniž by to bylo jakkoliv zdůvodněno. Zajímavé je, že hodnoty některých jiných parametrů byly vybrány jako ty s nejlepším výsledkem.
- Struktura práce je poněkud zmatečná na konci kapitoly 2 a v kapitole 3. Nejvíce matoucí je míchání principů s implementačními detaily (jména proměnných apod.). Zeména v kapitole 3, která by měla obsahovat těžiště přínosů práce, jsou tyto přínosy pouze načrtnuty v úvodních odstavcích a jsou následovány velmi technickým popisem jednotlivých proměnných a parametrů a s velkým důrazem na implementaci. Kvůli tomu je poměrně těžké vyextrahovat hlavní ideje a čtenář neznalý implementace a detailů může mít velké problémy porozumět hlavnímu sdělení práce.
- V kapitole 4 (Final experiments) jsou tvrzení, že algoritmus X je nejlepší, Y je nejhorší apod. Důkazy pro to jsou ale nedostatečné. Nejsou zde žádné statistické testy či znázornění rozptylu hodnot v grafech. U grafů dokonce ani není uvedeno, zda čáry reprezentují průměr či medián či jakoukoliv jinou hodnotu. Tvrdit, že něco je "nejlepší" není možné s tak chabými důkazy. Stejný problém je přítomen v závěrečné kapitole, kde jsou odpovědi na výzkumné otázky, které jsou založené na stejně špatnch důkazech - čárové grafy bez chybových úseček či statistických testů.
- První výzkumná otázka "Může CMA-ES zvládnout zašuměné funkce?" není, tak jak byla položena, ve skutečnosti zodpovězena, protože odpověď nemluví o CMA-ES ale o jeho rozšířených verzích. Tudíž otázka byla položena špatně nebo odpověď neodpovídá na otázku.
- Odpověď na druhou otázku ukazuje na problémy jí samotné, jelikož začíná slovy "it depends." Zbytek odpovědi také není příliš uspokojivý, jelikož neodkazuje na žádné důkazy a je to pouze vyjádření autorova názoru. Nicméně odpověď obsahuje důležité a hodnotné postřehy k tomu, jak se jednotlivé algoritmy liší (což ale mělo být uvedeno mnohem dříve u popisu samotných algoritmů) a jak by mohly být použity v součinnosti.
- Odpověď na třetí otázku začíná slovy "in my opinion." Podobně jako u předchozí otázky - na čem je tento názor založen?

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
5. Formální úroveň práce	*62 (D)*

Popis kritéria:
Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Komentář:
Angličtina, kterou je práce psaná, má problémy. V práci se vyskytují takové chyby, které by mohly být považovány za překlepy, ale vzhledem k jejich četnosti mám silný pocit, že se jedná čistě o chyby. Některé věty jsou velmi podivné, některé dokonce tak, že nedávají smysl nebo jsou velmi matoucí. To je velký problém v úvodní kapitole, kde je velmi důležité předat čtenáři hlavní cíle práce. V textu je mnoho typografických chyb (pravděpodoně se jedná o překlepy), jako např. závorky, které by tam neměly být, tečky po nichž nenásleduje mezera, znak rovnítka na konci řádku (a pravá strana výrazu následující na dalším řádku či v jednom případě dokonce na další stránce).

V hlavním textu práce je mnoho výpisů konfiguračních souborů. Tyto by měly být spíše v příloze nebo v elektronické formě. Text práce by měl parametry uvést vhodnější formou, např. tabulkami.

Použití matematických symbolů a konstrukcí je na velmi dobré úrovni.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Práce se zdroji	*51 (E)*

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:
Odkazovaná literatura vykazuje velkou míru nekvality. Například zcela chybí odkaz na původní článek/články o CMA-ES (např. Hansen, Nikolaus, and Andreas Ostermeier. "Completely derandomized self-adaptation in evolution strategies." Evolutionary computation 9.2 (2001): 159-195) a je zde odkaz pouze na nerecenzovaný tutorial na arxiv.org (nicméně od původního autora). Drtivá většina odkazované literatury jsou online zdroje, z nichž některé je možné považovat za nespolehlivé. Jedním z nich je i Wikipeda, kterou může upravovat kdokoliv, další je StackExchange, což je platforma pro pokládání a odpovídání otázek širokou veřejností, není to ověřený zdroj.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
7. Hodnocení výsledků, publikační výstupy a ocenění	*61 (D)*

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:
Práce prezentuje zlepšení oproti výchozím algoritmům, ze kterých práce vychází. Nicméně, kvůli tomu, jak jsou výsledky prezentované, je těžké usoudit skutečnou hodnotu těchto výsledků. Zdrojový kód vypadá čistě a bez větších problémů. Kód pracuje v kontextu většího frameworku a je (může být) jeho součástí.

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:
Práce se zbývá tématem black-box optimalizace. Algoritmy řešící tento úkol obecně mají uplatnění, takže práce na toto téma má téměř automaticky potenciál se uplatnit jak v praxi tak jako nástroj v dalších výzkumných oblastech.

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:
- Jak, ze statistického hlediska, ve skutečnosti vypadají výsledky, které byly v práci prezentované jako jednoduché čarové grafy? Nebo jednodušeji, jak vypadají chybové úsečky?
- Jak se v algoritmu evo-CMA-ES přepíná mezi evolucí modelů a vlastním CMA-ES a jak se rozhoduje, kolik času se alokuje pro který z těchto režimů?

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
10. Celkové hodnocení	*55 (E)*

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

Text hodnocení:
Práce zpracovává všechny body zadání a má solidní strukturu. Student na této práci vynaložil netriviální úsilí. Nicméně je zde několik problémů, které práci degradují. Dva nejzávažnější z nich jsou úroveň použité literatury a způsob, jakým jsou prezentovány a analyzovány výsledky. Ostatní problémy zahrnují jazykové chyby a překlepy a příliš technický popis přínosů namísto jasného a více srozumitelného a lépe pochopitelného. Práce působí dojmem, že byla napsána rychle a bez přílišné kontroly. Nicméně nápady v práci prezentované jsou zajímavé a z výsledků (i když jejich významnost je obtížné určit) se zdá, že tyto nápady fungují.

Celkově práci hodnotím jako ne příliš kvalitní, ale dostačující na stupeň E.

Podpis oponenta práce: