

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Student: Bc. Martin Hazda

Oponent: Ing. Lenka Skanderová, Ph.D.

Studijní program: **Inženýrská informatika**
Studijní obor/Specializace: **Kybernetická bezpečnost**
Akademický rok: **2021/2022**

Téma diplomové práce: **Nástroj pro identifikaci slabých míst evolučních algoritmů**

Hodnocení práce:

- **Úplnost vypracování, aktuálnost a obtížnost řešeného úkolu**

Práce si klade za cíl tvorbu nástroje pro identifikaci slabých míst evolučních algoritmů. Kromě samotného nástroje bylo úlohou studenta vypracovat literární rešerši, sestavit množinu vhodných testovacích problémů a navrhnout metodiku pro testování a vyhodnocení výsledků. Student se měl zaměřit na analýzu diverzity populace a vznik clusterů v populaci. Následně měl otestovat vybrané moderní evoluční algoritmy a zhodnotit získané poznatky.

Podle mého názoru se jedná o velmi obecné a poměrně obtížné zadání pro studenta magisterského studia vzhledem k tomu, že nelze předpokládat hluboký vhled do problematiky dynamiky evolučních algoritmů ani znalost moderních evolučních algoritmů, které v dnešní době oplývají nejrůznějšími „nadstavbami“.

Těžko se mi hodnotí splnění zadání, neboť metodika, kterou student navrhnul, stojí na poněkud vratkých základech, navíc student k otestování nepoužívá moderní evoluční algoritmy, kromě SOMA T3A, která byla představena v roce 2019. Dalšími algoritmy jsou kanonické verze diferenciální evoluce (DE, rok 1995) a optimalizace rojem částic (PSO, rok 1997), které bych jako moderní evoluční algoritmy neoznačila. U těchto algoritmů navíc není jasné, kdo je implementoval, neboť nejsou součástí příloh. Tedy výsledky algoritmů nelze ověřit. Otestování metodiky na třech jednoduchých algoritmech, z nichž dva rozhodně nespádají do moderních technik, a osmi triviálních funkcích mi nepřipadá adekvátní množství práce pro práci diplomovou.

- **Způsob a úroveň pojetí řešeného úkolu**

Metodika:

Student postavil svou metodiku na hodnocení čtyř kritérií – konvergence, diverzita populace, shlukování a velikosti pohybu jedinců po hyperploše. Z textové části práce mi není zcela jasné, jak hodnotí předčasnou konvergenci, neboť píše: „*Velikost zkoumané oblasti je určena na posledních 25% počtu vzorků konvergencí, které jsou ve smyčce ukládány do samostatných polí ke komparaci.*“ Znamená to tedy, že začíná sledovat předčasnou konvergenci až v poslední čtvrtině běhu algoritmu? Pokud ano, pak tento přístup neshledávám příliš šťastným, neboť předčasnou

konvergenci lze jistě detekovat v jakékoliv chvíli s využitím několika málo generací. Nehledě na to, že detekce předčasné konvergence až takto pozdě nemá příliš praktický význam.

V případě detekce shluků byl použit balíček `sklearn.cluster`, a sice metoda DBSCAN. Student už ovšem nezdůvodňuje, proč zrovna tento nástroj a z jakého důvodu volí řídicí parametry nástroje tak, jak je v práci popsáno. Např. parametr `EPS` je nastaven na 1% prohledávaného prostoru. Proč? `MinPts = 4`. Z jakého důvodu? Nastavení těchto parametrů je přitom pro algoritmus klíčové a podstatně ovlivní výsledek. Student se u tohoto nastavení odvolává na referenci [31]. To však není zdůvodnění.

Dynamika pohybu je pak postavena na Euklidovské vzdálenosti.

Čili lze konstatovat, že celá práce stojí na čtyřech jednoduchých vzorcích/popř. algoritmu, který poskytuje pythonovský balíček `sklearn.cluster`. To samo o sobě by nebylo až tak velkou překážkou, kdyby byla navržená metodika řádně otestována, student vzal do úvahy i jiné metody, které by bylo možné použít, a dal si práci aspoň např. s implementací moderních evolučních algoritmů a moderních benchmarků.

Sestavení množiny vhodných testovacích problémů:

Student věnuje tzv. benchmarkům celou Kapitolu 3, kde zmiňuje např. benchmark IEEE CEC 2022 a dokonce uvádí i novější verze evolučních algoritmů. Určitě bych je nenazývala algoritmy soudobými, vzhledem k tomu, že jDE, SaDE, ARPSO atd. jsou algoritmy poměrně staršího data, např. ARPSO bylo představeno v roce 2002. Zajímavé je, že student činí rozdíly mezi soudobými algoritmy a state-of-the-art algoritmy. Není ovšem jasné, podle jakého kritéria algoritmy do těchto skupin dělí.

Kuriózní na celé Kapitole 3 je fakt, že je de facto zbytečná. V praktické části pro otestování není použita ani jedna funkce z moderních benchmarků a ani jeden z tzv. soudobých algoritmů.

Testování algoritmů:

Student si pro testování zvolil pouhých 10 dimenzí a počet jedinců nastavil pro všechny algoritmy na hodnotu 50. Rojové inteligenci však velký počet jedinců vyloženě škodí. Tedy nastavení takto velké populace pro takto malou dimenzi problému mi připadá nešťastné. Nastavení není řádně zdůvodněno.

Zajímavé je, že student si jako testovací funkce určil 8 triviálních funkcí (o smysluplnosti použití některých z nich pochybuji, např. konstantní funkce), ale závěry tvoří pouze na základě dvou. Velmi zajímavý je i výsledek diferenciální evoluce pro Schwefelovu funkci (strana 72), kde studentovi vychází hodnota účelové funkce záporná a velmi blízko 0, a on konstatuje, že v tomto případě došlo k aproximaci globálního optima. Což ale podle toho, jak definuje Schwefelovu funkci není pravda, protože podle jeho definice v sekci 5.5.3 je hodnota účelové funkce v globálním extrému rovna hodnotě $-418.983 \cdot D$ (dimenze). Samozřejmě lze definovat Schwefelovu funkci tak, aby měla globální extrém s hodnotou účelové funkce rovnou 0, ale v takovém případě není možné, aby hodnota účelové funkce nabývala záporných hodnot.

Závěr, který říká, že ve verzi DE/rand/1/bin převažuje exploitace a že ztráta diverzity a vznik shluku vedly k aproximaci globálního extrému, je chybný. Myslím, že si student plete příčinu a

následek. Nehledě na to, že v tomto případě se jedná o přirozenou konvergenci, a tím pádem i o přirozenou postupnou ztrátu diverzity, což ukazuje i studentem uváděný graf.

Aplikace:

Aplikace sestává ze 4 pythonovských souborů. Dva soubory se starají o práci s csv soubory, pak je zde jeden soubor pro analýzu a poslední pro spuštění aplikace. Množství kódu a jeho kvalita je na diplomovou práci a studenta magisterského studia tristní. Student nevynaložil úsilí na implementaci moderního benchmarku, ani moderních evolučních či rojových algoritmů. V souvislosti s množstvím provedených testů nelze praktickou část práce hodnotit ani jako dostatečnou.

- **Úroveň zpracování tématu, přínos diplomanta**

Úroveň zpracování tématu jak po stránce implementační, tak po stránce experimentální a textové hodnotím jako velmi nízkou. Práce na mě působí jako cesta nejmenšího odporu, kdy student vybral tři jednoduché algoritmy, velmi jednoduché testovací funkce a čtyři metriky, na nichž založil metodiku pro identifikaci slabých míst evolučních algoritmů. Následně popsal otestování vybraných algoritmů na pouhých dvou z osmi funkcí.

- **Formální náležitosti práce, chyby a omyly v technické zprávě**

Úroveň textové části práce není valná. Student ignoruje existenci odstavců, obrázky a tabulky jsou v neodpovídající kvalitě. Text je velmi rozvláčný a jak již bylo zmíněno, některé kapitoly jsou zcela zbytečné vzhledem k následné praktické části.

Některá tvrzení jsou více než sporná, ba přímo nepravdivá např.: „*Optimum reálného problému lze jen obtížně nalézt analytickými nebo numerickými metodami s ohledem na výpočetní náročnost nebo proveditelnost...*“ Reálné problémy se běžně numerickými a analytickými metodami řeší. A jejich užití je mnohem častější než užití evolučních technik. Student také např. tvrdí, že: „*EA najdou uplatnění například při trénování neuronových sítí...*“ EA se však pro trénování neuronových sítí nepoužívají a těžko zde najdou uplatnění ve stejné míře jako osvědčené metody založené na gradientu.

Účelová funkce není totéž co tzv. cenová (cost) funkce. Cenová funkce se používá výhradně ve spojení s minimalizací.

- **Dotazy k obhajobě**

Jakým způsobem se vyhodnocuje předčasná konvergence?

Z jakého důvodu jsou v práci použity tak jednoduché funkce, když v Kapitole 3 popisujete nový benchmark IEEE CEC 2022. Existují benchmarky typu CEC i z minulých let. Proč jste nepoužil některý z nich?

Jaký smysl dává konstantní funkce pro otestování evolučních algoritmů?

V sekci 3.3 popisujete Soudobé algoritmy, ale v práci jste je nevyužil. Totéž platí i pro sekci 3.4 State of the art algoritmy. Z jakého důvodu, když ze zadání vyplývá, že se máte soustředit na moderní algoritmy?

Jaký je podle Vás rozdíl mezi soudobými algoritmy a state-of-the-art algoritmy? Proč je např. SaDE zařazeno do soudobých algoritmů a L-SHADE do state-of-the-art algoritmů?

Kdo implementoval algoritmy DE, PSO a SOMA T3A?

Jak byla definována Schwefelova funkce? Vysvětlete výše popsany rozpor ohledně výsledků u diferenciální evoluce.

Proč nejsou v práci uvedeny a vyhodnoceny výsledky na všech vybraných funkcích?

- **V závěru zhodnoťte celkově předloženou diplomovou práci a klasifikujte dle klasifikační stupnice uvedené v závěru tohoto formuláře.**

Práce se mi obecně velmi špatně hodnotí. Na jedné straně konstatuji, že zadání je pro studenta magisterského studia příliš obecné a poměrně obtížné. Na druhou stranu toto zadání bylo schváleno a student si jej pravděpodobně dobrovolně vybral. Čili lze předpokládat, že se mu bude systematicky, zodpovědně a pečlivě věnovat.

Bod zadání, který říká, že má student k testování vybrat moderní algoritmy, nepovažuji za splněný. Kromě SOMA T3A byly vybrány původní verze dobře známých technik (DE a PSO), které za moderní nelze považovat. Navíc testování celé metodiky na základě pouhých tří algoritmů a jedním nastavením řídicích parametrů nelze považovat za dostatečné.

Zvolené funkce pro testování jsou triviální (některé z nich nedávají pro otestování takových algoritmů smysl), nemají nic společného s moderními benchmarky, a student jejich použití v práci dostatečně nezduvodnil. Výběr těchto funkcí je navíc tím nepochopitelnější, že v Kapitole 3 popisuje nejmodernější benchmark typu CEC.

Metodika, kterou student zvolil, je založena na čtyřech kritériích jednoduchého charakteru. To samo o sobě není zase až takovou překážkou, nicméně tato metodika není dostatečně zdůvodněná (viz poznámky výše) a hlavně otestovaná. Testování postrádá jakýkoliv systém, statistické testy, zkratka cokoliv, na čem by bylo možné postavit závěr, že je daná metodika korektní.

Textová část práce obsahuje velké množství textu, který přímo s jádrem zadání práce nesouvisí, tedy nepřispívá k jeho naplnění. Nemá význam v práci popisovat nové benchmarky včetně způsobů jejich vyhodnocení a novější verze algoritmů, které nakonec nejsou využity. Rovněž nemá význam rozvláčně popisovat tak jednoduché funkce, které byly v práci použity. Obrázky těchto funkcí jsou rovněž zbytečné. Větší pozornost měla být soustředěna na výběr kritérií pro sestavovanou metodiku, hypotézám, testům, výběru a implementaci algoritmů pro testy, nastavení jejich řídicích parametrů a výběru a implementaci vhodných benchmarků.

V současném stavu hodnotím práci jako nedostatečnou a nedoporučuji ji k obhajobě. Doporučuji práci dopracovat dle poznámek popsanych výše.

Celkové hodnocení práce:

Známku uvede oponent dle svého uvážení dle klasifikační stupnice ECTS:

A – výborně, B – velmi dobře, C – dobře, D – uspokojivě, E – dostatečně, F – nedostatečně.

Stupeň F znamená též „nedoporučuji práci k obhajobě“.

Předloženou diplomovou práci nedoporučuji k obhajobě a navrhuji hodnocení

F - nedostatečně.

V případě hodnocení stupněm „F – nedostatečně“ uveďte do připomínek a slovního vyjádření hlavní nedostatky práce a důvody tohoto hodnocení.

Datum 1. 6. 2022

Podpis oponenta diplomové práce