

## Oponentský posudek doktorské disertační práce

### Optimalizační metody pro vícevrstvý algoritmus genetického programování

Doktorand: Ing. Jan Merta

Školitel: doc. Dr. Ing. Tomáš Brandejský

Oponent: doc. Ing. Vít Fábera, Ph.D., Ústav aplikované informatiky v dopravě, FD ČVUT

Předložená disertační práce je poměrně obsáhlá, celkem 103 stran, včetně seznamu prací doktoranda. Je rozdělena do 9 kapitol (desátá je přehled literatury), přičemž těžiště práce začíná pátou kapitolou, kde jsou uvedeny výsledky dosavadní původní práce doktoranda.

Doktorand se ve své práci zaměřuje na optimalizaci algoritmu genetického programování, konkrétně vylepšení algoritmů genetického programování pro oblast symbolické regrese. Téma, které doktorand zvolil, je aktuální a potřebné, protože neustále roste aplikační oblast, kde se tyto techniky uplatňují, a zejména množství a složitost modelů, pro jejichž úspěšné zpracování přestávají algoritmy ve své původní podobě stačit – ať již dobou výpočtu či úspěšností měřenou zpravidla nějakou chybovou funkcí.

Úvodní rešeršní kapitoly mají široký záběr. Kromě nezbytného stručného popisu principů genetického programování a souvisejících problémů s ním (na úrovni dostatečné pro disertační práci, nikoliv zbytečně podrobné) doktorand uvádí další optimalizační metody používané ve strojovém učení a vylepšení genetických algoritmů: redukce dimenzionality, sborové učení (ensemble learning), hybridní a paralelní genetické algoritmy. Z textu patrné, že doktorand má skutečně široký přehled o dané problematice a prostudoval patřičnou literaturu, což dokazuje její seznam čítající 69 položek, přesto bych uvítal ve výčtu ostatních metod např. také diferenciální evoluci. Negativum širokého přehledu vidím snad jedině v tom, že čtenáři se na první pohled může zdát, že autor popisuje několik spolu nesouvisejících metod a není zcela jasné, jakým konkrétním směrem se bude autorův přístup dále rozvíjet.

Vlastním, původním přínosem je dvouvrstvý algoritmus genetického programování. Od existujících přístupů, kdy jiná vrstva optimalizuje zpravidla pouze konstanty, doktorand navrhl „opačný“ přístup, tj. stromy včetně konstant vytváří již na první vrstvě a druhá vrstva skládá evolučně modely vytvořené v první vrstvě (inspirováno sborovým učením). Oceňuji celkový přístup autora k problematice. Využil svých předchozích prací (např. porovnání lamarckistického a baldwinistického přístupu pro optimalizaci konstant, na základě jehož výsledků vybírá lamarckistický přístup nebo zkušenost s algoritmem Bison Seeker) a využil je při návrhu dvouvrstvého algoritmu. Dvouvrstvý algoritmus by si zasloužil v práci podrobnější popis, k čemuž směřuje moje první otázka.

Příklady jsou voleny vhodně: najít regresní model analytické funkce hyperbola, sinus, lineární kombinace sinů a reálných dat průběhu teploty v indickém městě Dillí. Autor záměrně (a přirozeně) vynechává z množiny neterminálů funkce, které nebyly u prvních tří příkladů použity. Srovnává jednovrstvý algoritmus s dvouvrstvým různě parametrizovaným (způsobu výběru trénovacích dat, zakomponování BSA, různé množiny funkcí co by neterminály na obou vrstvách).

Ke způsobu uvádění výsledků nemám téměř připomínky. Parametrizace algoritmů je podrobně popsána, experimenty statisticky vyhodnoceny a zobrazeny graficky (průměrné hodnoty, rozptyly, odlehle hodnoty) a ukazují, že autorův přístup zřejmě vede k uspokojivým

výsledkům - dalšímu vylepšení algoritmů genetického programování aplikované na problematiku regrese. Škoda jen, že autor neuvedl u jednotlivých běhů algoritmů ukázky výrazů, které algoritmy našly.

Seznam literatury považuji za odpovídající úrovni práce, jsou zde citovány nové publikace (z roku 2021). Seznam vlastních prací doktoranda obsahuje 18 položek, jeden ve vědeckém časopisu.

Práce je po formální stránce napsaná přehledně a čtivě, na velmi dobré odborné i jazykové úrovni, obsahující minimum překlepů (našel jsem snad jednu méně srozumitelnou formulaci).

Ve své doktorské práci prokázal Ing. Jan Merta schopnost samostatné vědecké práce. I přes drobné nedostatky splňuje předložená práce všechna potřebná ustanovení pro udělení titulu „Ph.D.“ dle § 72 odst. 5 zákona č. 111/1998 Sb. o vysokých školách a doporučuji ji k obhajobě.

K práci mám následující dotazy:

1. Můžete popsat, jak jsou vybírány dílčí modely vzniklé v první vrstvě do druhé vrstvy?
2. Uvažoval jste o restartu algoritmu v následujícím smyslu: pokud by došlo k dlouhodobé stagnaci fitness funkce ve druhé vrstvě, spustit opětovně první vrstvu algoritmu, event. při opakovaném běhu první vrstvy využít informace z vrstvy druhé?
3. Dovedete odhadnout, zda je Váš dvouvrstvý algoritmus díky druhé vrstvě zvlášť vhodný na průběhy, resp. funkce, které se dají vyjádřit vhodně jako řada, nebo nemá žádnou výhodu, resp. za jakých podmínek?
4. Můžete uvést příklady nalezených funkcí genetickým programováním? Odpovídá u funkce sinus třeba řadě  $\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}$ ?
5. Jak se díváte na využití diferenciální evoluce pro optimalizaci konstant v další vrstvě algoritmu?

doc. Ing. Vít Fábera, Ph.D.

Ústav aplikované informatiky v dopravě

Fakulta dopravní ČVUT v Praze

V Praze dne 19. 08. 2022