

Oponentní posudek disertační práce Ing. Tomáše Urbánka „*Využití analytického programování pro odhady časových náročností vývoje softwarových projektů*“

Ing. Tomáš Urbánek se ve své disertační práci zabývá návrhem a implementací frameworku pro odhad časové náročnosti vývoje softwarových projektů, čímž tematicky navazuje na habilitační práce doc. Petra Šilhavého a doc. Radka Šilhavého, z nichž druhý je současně jeho školitelem.

Obě habilitační práce ukázaly, jak odlišnými směry lze k řešení takových odhadů přistupovat, zatímco doc. Petr Šilhavý k odhadu pracnosti využívá *funkční body* (základní charakteristiky vyvíjeného softwaru, datové funkce, transakční funkce a zohledňuje jejich složitosti, související s počtem externích vstupů a výstupů, počtem externích dotazů, počtem vnitřních logických souborů a počtem souborů vnějších rozhraní), z nichž se určí váhový součet a po následné modifikaci multiplikativními faktory přizpůsobení a produktivity se z upravených funkčních celků stanoví odhad pracnosti v člověkohodinách, doc. Radek Šilhavý se zabývá určováním *rozsahu softwarových systémů*, tedy počtu funkcí, které systém bude zajišťovat a které (do určité míry) jsou rovněž úměrné náročnosti vývoje softwaru.

Na druhé straně zmíněné charakteristiky a jejich kvantitativní vyhodnocení nemusí dávat úplně přesný obraz náročnosti vývoje softwaru, protože sw aplikace pro vědeckotechnické výpočty (např. pro řešení nelineárních diferenciálních rovnic), byť s malým počtem vstupů, mohou být mnohem obtížnější než databázové aplikace pracující se stovkami tisíc záznamů a jednoduchými operacemi vložení, opravy a rušení dat a synchronizace transakcí. Kód softwaru může být unikátní pro daný účel, anebo také do značné míry daný „koláží“ knihovnicích modulů z volně dostupných zdrojů či opakovaných, popř. jen mírně modifikovaných verzí dřívějších řešení.

Řadu problémově orientovaných softwarových aplikací lze také implementovat odlišnými postupy a určující je funkčnost softwaru a ne přesná podoba zdrojového kódu (viz třeba velký počet různých programů pro účetnictví).

Vzhledem k velkému počtu těžce postižitelných faktorů i míře neurčitosti je stanovený úkol netriviální a při zvládnutí cíle je práce **disertabilní**.

Hlavní cíl disertace autor v 2. kapitole upřesňuje specifikací mnoha postupných úkolů, které povedou k řešení. Vedle standardních, jako jsou rešerše existující literatury z oblasti výzkumu a získání dat pro experimentální ověření, klade důraz na eliminaci vlivu lidského činitele, testování, zda existuje lineární závislost odhadu náročnosti navrženým frameworkem na skutečném čase vývoje, a samozřejmě také vyhodnocení přesnosti odhadu při variantně volených účelových funkcích a porovnání s výsledky získanými jinými metodami a softwarovými nástroji (např. různými typy regrese a implementací vícevrstvé neuronové sítě či rozhodovacího stromu).

Důležitou částí práce je teoretický rozbor metod softwarového inženýrství, jejich parametrů a postupů nastavení hodnot či vah. Blíže je popsána metoda Use Case Points, založená (jak říká i její název) na případech užití. Autor však dává přednost symbolické regresi a z metod pro nalezení regresní funkce, kam patří genetické programování, gramatická evoluce a analytické programování, volí poslední metodu, protože nevykazuje nečtnosti ostatních, např. efekt „bloat“ (bobtnání řešení) u genetického programování a obtíže u gramatické evoluce při určení gramatiky, specifické pro programovací jazyk, v němž je projekt implementován.

K vyhodnocení kvality modelu softwarového inženýrství je potřebné stanovit jejich kritérium, a tím se autor zabývá v odstavci 4.3, kde uvádí analytická vyjádření nejpoužívanějších kritérií MMRE (Mean Magnitude of Relative Error), LAD (Least Absolute Deviation) a MSE (Mean Squared Error).

Stěžejní částí disertační práce jsou v Kapitolech 6 a 7. V 6. kapitole autor popisuje navržený framework, použité datasety a statistické testy, které bude k potvrzení hypotéz využívat, a návrh experimentů. Kapitola 7 pak dosažené výsledky podrobně diskutuje a statisticky vyhodnocuje (např. pomocí testů Shapiro-Wilk, ANOVA a Kruskal-Wallis). Výsledky jsou popsány v tabulkách a pro názornost prezentovány i graficky.

Disertační práci završují shrnující a hodnotící Kapitoly 8 a 10, mezi něž je zařazena stručná Kapitola 9, vytyčující možné směry dalšího výzkumu. I z toho je zřejmé, že nelze očekávat jednoznačný a jediný možný postup řešení odhadů náročnosti vývoje softwaru a jak autor uvádí, přesnost různých přístupů závisí i na velikosti datasetů. I tak se však autorovi podařilo prokázat, že jím navržený framework na většině testovaných dat prokazuje vyšší přesnost než známé a standardně používané metody. A to je největším přínosem disertační práce.

Formální úroveň práce je velmi dobrá, snad jen pojmy „komplexní“ a „komplexita“ jsou anglicismy, v českém textu by bylo vhodnější využít česká slova „složitý“, „složitost“. Např. u *time complexity* mluvíme o časové složitosti, nikoliv časové komplexitě.

Vhodné by také bylo zaváděné pojmy v prvním výskytu psát pro větší čitelnost kurzívou. Symboly v matematických vztazích psané kurzívou by takto měly být uváděny i v textu, např. na str. 22, 23 a 24 místo „í-tého“ by mělo být „*i*-tého“ (krátké *i* a kurzíva) a na str. 25 není dodrženo u *EE*, *UCP*, *FP*.

Při uvádění síly testu autor střídavě používá vyjádření desetinnými čísly menšími než 1 a na jiných místech v desítkách procent, a to i na sousedních stránkách (např. 77 a 78), pro jednotnost výkladu by bylo lepší vybrat jednu možnost.

Jazyková úroveň práce je však mnohem slabší, především velmi často chybí čárky ve složitějších souvětích, někdy jsou na nevhodných místech, vyskytují se ale i jiné chyby. Následující výčet není úplný.

V poděkování za čas a rady doc. Prokopové a doc. Šilhavému má být „věnovali“ a „poskytli“ (a ne „věnovaly“ a „poskytly“).

Str. 13: „Má však také velké sociální dopady jak dokládá studie“ – chybí čárka před „jak“.

Str. 13: „Mnoho produktu, které byly tradičně realizovány formou hardwaru jsou dnes řešeny formou softwaru“ – chybí čárka za slovem „hardwaru“, kde končí vložená vedlejší věta.

Str. 14: „zahrnuje spoustu faktorů jako je“ – chybí čárka před „jako“.

Str. 17: „dala základy toho jak a kdy je nutné ohodnocovat“ – chybí čárka před „jak“.

Str. 18: „kdy se dozvíme jakým způsobem se dospělo k výsledku“ – chybí čárka před „jakým“.

Str. 18: „problematikou 'bloat' tedy bobtnání řešení“ – chybí čárka před „tedy“.

Str. 18: „se snaží vylepšit ... metody jako je např. metoda“ – chybí čárka před „jako“.

Str. 19: „srovnání podporujících technik ..., kdy dospěly k názoru“ – co je podmětem? Techniky? Ty asi nemají názor. Pokud jsou myšleni autoři, pak „dospěly“ může být jen v případě, když jde výhradně o ženy.

Str. 20: „metody softcomputingu např. Štrba et al.“ – patří čárka před „např.“

Str. 20: „když nemáme k dispozici historické projekty není možné zkonstruovat odhad“ – chybí čárka za slovem „projekty“.

- Str. 21: „zařadit do jedné ze zmíněných kategorií a to nejčastěji metody řízené daty“ – chybí čárka před „a to“.
- Str. 21: „Proto, popsanou metodu v této práci řadíme do metod řízených daty.“ – čárka sem nepatří.
- Str. 24: „konstanta, která udává kolik je potřeba“ – chybí čárka před „kolik“.
- Str. 26: „Efekt "bloat" vzniká když začne být“ – chybí čárka před „když“.
- Str. 27: „To znamená že např. funkce ... obsahuje“ – chybí čárka před „že“.
- Str. 29: „kolik proběhne iterací algoritmu než bude ukončen“ – chybí čárka před „než“.
- Str. 31: „**B** je matice koeficientu, které budou odhadovány a **U** je matice chyb“ – chybí čárka za slovem „odhadovány“.
- Str. 31: „Postupy, které se u krokové regrese uplatňují jsou následující“ – chybí čárka za slovem „uplatňují“.
- Str. 32: „Vyřazování proměnných končí až jsou“ – chybí čárka před „až“.
- Str. 32: „tyto porovnání“ – má být „tato porovnání“.
- Str. 33: „Ridge regrese někde také česky označována jako hřebenová regrese je metoda“ – „Ridge regrese, někde také česky označována jako hřebenová regrese, je metoda“.
- Str. 33: „může snížit rozptyl reziduí a tím zpřesnit odhad“ – chybí čárka před „a tím“.
- Str. 34: „Pokud je k nastaveno na hodnotu 1 pak je bod klasifikován do stejné třídy jako má nejbližší objekt“ – chybí čárky před slovy „pak“ a „jako“.
- Str. 35: „Algoritmus funguje ... a zároveň se snaží aby tato hranice“ – chybí čárka před „aby“.
- Str. 35: „Data, pro která získáváme predikci jsou potom zařazena“ – chybí čárka za slovem „predikci“.
- Str. 35: „která určuje jak moc bude signál“ – chybí čárka před „jak“, „moc“ je hovorové slovo.
- Str. 35: „Jak moc budou přenastaveny váhy je funkcí rozdílu“ – chybí čárka před „je“.
- Str. 36: Přístupy, které byly použity v tomto výzkumu jsou uvedeny – chybí čárka za slovem „použity“.
- Str. 37: „Když je $k = 1$ tak má“ – chybí čárka před „tak“.
- Str. 37: „Mnoho vědeckých studií se zabývá také otázkou jaké zvolit kritérium“ – chybí čárka před „jaké“.
- Str. 38: „ačkoliv je MMRE kritizováno je využíván“ – chybí čárka za slovem „kritizováno“.
- Str. 40: „Toto se děje z důvodu použití druhé mocniny ve výpočtu a tedy odlehle hodnoty jsou umocněny“ – chybí čárka před „a tedy“.
- Str. 40: „výsledkem je spojitá funkce konkrétně parabola“ – za slovem funkce by měl být oddělovač (čárka, pomlčka, dvojtečka).
- Str. 41: „[? 74]“ – chybný kód citace.
- Str. 42: „budou podrobeny klasickým metodám matematické statistiky jako je deskriptivní (popisná) statistika“ – chybí čárka před „jako je“.
- Str. 42: „z veřejně dostupných databází popř. prostudováním akademických studií“ – chybí čárka před „popř.“.
- Str. 43: „Tímto by jsme získali“ – „... **bychom** ...“

Str. 43: „podle Karneroví rovnice“ – „... Karneroví ...“

Str. 44: „pomocí stratifikovaného případně vícestupňového výběru dat“ – před „případně“ patří čárka.

Str. 44: „parametry metody UCP tzn. UUCW“ – před „tzn.“ patří čárka.

Str. 44: „považován za historický (nese novou informaci) je uložen do databázového systému, a může být“ – před „je uložen“ patří čárka (anebo spojka „a“), naopak před „a může“ čárka nepatří.

Str. 45: „byly datasety používány zvlášť v pořadí v jakém docházelo“ – chybí čárka před „v jakém“.

Str. 46: „sestavá z 10-ti softwarových projektů“ – „... 10 ...“

Str. 46: „je k dispozici 24 záznamů a to znamená“ – chybí čárka před „a to“.

Str. 47: „se průměr UUCW zvýšil o 2,7 krát“ – „o“ sem patří.

Str. 47: „Co se týká výpočtu směrodatných odchylek jednotlivých proměnných (sd) ty jsou“ – chybí čárka za „(sd)“.

Str. 47: „obsahuje více pozorování a tedy se rozptýl“ – chybí čárka před „a to“.

Str. 47: „Pearsonoví korelační koeficienty“ - „Pearsonoví...“, na stránce se vyskytuje 2krát.

Str. 47: „Žádný z parametru nevykazuje vizuální známky normálního rozdělení snad jen parametr ECF“ – patří čárka před „snad“.

Str. 47: „lognormální rozdělení popř. jiné rozdělení“ – před „popř.“ patří čárka.

Str. 50: „Avšak pro dataset D2, se podařilo nasbírat“ – čárka sem nepatří.

Str. 50: „pro oba datasety vyšly, tak diametrálně rozdílné výsledky“ – čárka sem nepatří.

Str. 51: „skončíme v minimu i když“ – před „i když“ má být čárka.

Str. 51: „obsahovala třináct symbolů viz. rovnice 6.1“ – před „viz“ má být čárka, totéž na str. 52.

Str. 52: „hypotéza, že by jsme neměli“ – „... bychom ...“

Str. 52: „Aby mohla být lineární regrese použita je zapotřebí“ – chybí čárka za slovem „použita“.

Str. 52, 54, 56, 57: „využity klasické hodnoty tedy 66 % dat bude použito“ – před „tedy“ patří čárka.

Str. 53: „Méně než v 50ti % případech“ – „... 50 % ...“

Str. 54: „může dojít ke zjednodušení odhadovacího modelu resp., může dojít“ – čárka má být před „resp.“.

Str. 54: „bude sestaven následující experiment viz. obrázek“ – před „viz.“ patří čárka.

Str. 54, 57, 60, 69, 77: „ $1 - \beta = 0,8$ “ – „0,8“ (bez mezery mezi číslicemi, tj. 8 desetin).

Str. 56: „řešena pomocí prezentovaného frameworku tedy pomocí“ – před „tedy“ patří čárka.

Str. 57: „Pro pilotních 10 výběrů budou sestaveny dva modely jeden podle“ – před „jeden“ patří čárka.

Str. 61, 62: „ $1 - \beta = 80\%$ “ – bylo cílem vyjádřit „osmdesátiprocentní“ (tj. 80%), anebo „osmdesát procent“ (80 %)?

Str. 61: „je zapotřebí 27 vzorků aby byla“ – chybí čárka před „aby“.

Str. 61: „ $1 - \beta = 81, 18\%$ “ – má být „81,18“ (bez mezery).

Str. 61: „vysvětlí průměrně 66 % variability resp. 30 % variability . Což je velmi nízká hodnota pro odhadovací model.“ – před „resp.“ patří čárka; před tečku nepatří mezer; věta nemůže být tvořena pouze vedlejší větou (místo tečky by mohla být čárka a navázáno s malým písmenem v podobě „což“).

Str. 63: “(standardní odhadovací metoda mnohonásobné lineární regrese)“ – závorky se píší těsně (bez oddělení mezerou).

Str. 63: „taky“ – „také“.

Str. 64: „Jelikož máme 15 skupin pro porovnání je potřeba“ – chybí čárka za slovem „skupin“ (anebo před „je potřeba“).

Str. 64-65: „První rovnice, které měly přístup k parametru UUCW si v konečném hodnocení vedly“ – chybí čárka za „UUCW“

Str. 68: „není ve skupině s žádným jiným parametrem a tedy je“ – chybí čárka před „a tedy“.

Str. 68, 72: „Hlavní motivací, řešení“ – čárka sem nepatří.

Str. 68: „některý s parametrů projektu“ – „některý z parametrů ...“

Str. 69: „Čím více máme projektů na základě nichž dochází ke generování rovnic tím menší“ – chybí čárka před „na základě nichž“ a za slovem „rovnice“.

Str. 69: „Jelikož máme 3 skupiny (MSE, LAD, MMRE) pro porovnání je potřeba“ – chybí čárka za pravou závorkou (anebo před „je potřeba“).

Str. 69: „Účelové funkce MMRE a LAD si vedla“ – „vedly“?

Str. 73: „10 průchodů, aby jsme mohly spočítat“ – „... abychom ... mohli ...“

Str. 73, 74, 85: „cca.“ – tečka za „cca“ nepatří.

Str. 73: „se průměr pro framework pohybuje na 8,5 % MMRE pro Karnerovu rovnici“ – za zkratkou má být čárka (a stejně tak na str. 74 v 2. řádku shora).

Str. 74: „Pro ověření, zda prezentovaný framework statisticky významně zpřesňuje odhad bude použit“ – chybí čárka za slovem „odhad“

Str. 74: „Bude použit párový t-test jelikož“ – chybí čárka před „jelikož“.

Str. 75: „Toto platí pro oba datasety tedy D1 i D2.“ – chybí čárka před „tedy“.

Str. 81: „Tyto modely, jak bylo ukázáno na grafu 7.11 měly horší statistické vlastnosti“ – chybí čárka za „7.11“.

Str. 84: získání nejpřesnějších výsledků tzn. nezajímají nás“ – před. „tzn.“ patří čárka.

Str. 84: „Avšak, jakmile začneme zkoumat jejich statistické vlastnosti zejména splnění Gauss-Markovových podmínek, potom cca. 30 % výběrů na datasetech nesplňovalo tyto podmínky a tedy použití lineární regrese“ – před „jakmile“ čárka být nemá; před „zejména“ chybí; „cca.“ bez tečky; před „a tedy“ má být čárka.

Str. 84: „Jak bylo ukázáno ve výzkumné otázce 2 nejdůležitějším parametrem je UUCW.“ – chybí čárka za číslovkou „2“.

Str. 84: „Toto se ukázalo, ale jen na menším datasetu D1.“ – před „ale“ čárka nepatří.

Str. 84: „Důvod, proč toto nebylo ukázáno na datasetu D2 je zřejmě ten“ – chybí čárka za „D2“.

Str. 84: „Ve výzkumné otázce 3, bylo ukázáno“ – čárka sem nepatří.

Str. 86: „jsou k dispozici doporučení jak velký by měl být poměr“ – chybí čárka před „jak“.

Str. 86: „tyto doporučení“ – má být „tato doporučení“.

Str. 86: „je uvedena komparativní studie jejíž závěry jsou“ – chybí čárka před „jejíž“

Str. 87: „jak analytického programování tak diferenciální evoluce“ – chybí čárka před „tak“.

Str. 87: „zkoumán a analyzován viz práce [70]“ – před „viz“ patří čárka.

Dotazy na doktoranda:

1. Jakým způsobem se vybírají rodiče pro generování jedinců do další nové populace (viz výklad na str. 30)?
2. Jaké softwarové nástroje jsou využity pro statistické testy?
3. Bude se navržený framework pro odhadování náročnosti v praxi využívat? A pokud ano, pak jak? Pro stanovení smluvní ceny či odměny, termínu dokončení projektu?
4. Všechny Vaše publikace jsou kolektivním dílem nejméně tří autorů. Jaký je Váš podíl na jejich výsledcích?

Závěr:

Disertační práce Ing. Tomáše Urbánka prokázala tvůrčí schopnosti autora navrhnout a implementovat efektivní nástroj pro odhadování náročnosti softwarových projektů a statisticky vyhodnotit a interpretovat validitu získaných kvantitativních výsledků. Doktorand do této chvíle je publikoval již v 17 článcích v časopisech (zvláště cenná je publikace v impaktovaném časopise *Applied Sciences*), kapitolách knih a sbornících konferencí, vesměs zařazených do vědeckých databází Web of Science a SCOPUS. Výrazně tak překročil požadovaná kritéria k úspěšnému ukončení doktorského studia. Práce je významná jak z pohledu teoretického přínosu k rozvoji oboru v oblasti softwarového inženýrství, tak i v praxi sw firem.

Domnívám se ale, že vzhledem k značnému množství gramatických chyb by alespoň pro účely elektronické archivace disertačních prací bylo vhodné text opravit. Po odborné stránce je však velmi přínosný a splňuje nároky vědecké práce, a proto

disertační práci Ing. Tomáše Urbánka

doporučuji k obhajobě

před komisí doktorského studijního oboru Inženýrská informatika

V Brně dne 24. listopadu 2020



Prof. RNDr. Ing. Miloš Šeda, Ph.D.
Ústav automatizace a informatiky
Fakulta strojního inženýrství VUT v Brně

Oponentský posudek doktorské disertační práce

Autor doktorské práce:

Ing. Tomáš Urbánek

Téma doktorské práce:

Využití analytického programování pro odhady časové náročnosti vývoje softwarových projektů

Oponent doktorské práce:

Doc. RNDr. PaedDr. Hashim Habiballa, Ph.D., Ostravská Univerzita

Téma práce a splnění cíle

Odhadování časové náročnosti vývoje projektů je složitou úlohou softwarového inženýrství, pracující s mnoha vstupními parametry a v současnosti je řešen klasickými metodami s poměrně velkou chybovostí. Zároveň to ale má významné ekonomické dopady pro firmy zabývající se vývojem softwarových produktů a správný odhad, může uspořit jak finanční, tak lidské zdroje. Autor vycházel jednak z klasických přístupů pomocí UCP (Use Case Points), které poskytuje odhad parametrů projektů v rané fázi vývoje a autor dále navrhl, implementoval a ověřil účinnost vlastního frameworku, schopného poskytnout lepší (přesnější) odhady. Z tohoto pohledu je práce aktuální a přínosná nejen pro poznání, ale i pro praxi.

Práce má jednak teoretickou a jednak praktickou (aplikační) část. Teoretická část začíná popisem metod UCP a nastavení jejich vah pro jednotlivé případů užití a aktérů, dále navazují popisy faktorů (technický, prostředí, produktivity). Dále se zabývá jednotlivými metodami použitými v originálním hybridním frameworku, zejména symbolické regrese, analytickému programování a diferenciální evoluci využité pro optimalizaci regrese pro výpočet odhadů časových náročností. Také je zde krátce rozebrána problematika neuronových sítí a jejich učení. Nakonec je v teoretické části rozebrána škála možných kritérií kvality (MMRE, LAD, MSE). Experimentální část pak popisuje tvorbu frameworku pro odhadování UCP s využitím parametrů UUCW, UAW, TCF a ECF. Framework pak s pomocí DE a analytického programování vytvoří rovnici pro odhad, která se aplikuje na základě parametrů projektu na konkrétní případ. Tento framework je pak podroben experimentům na účinnost navržené metody a to na 2 datasetech (od jiných autorů) – s 24 a 70 projekty.

Cíle zahrnují:

- Návrh frameworku k získání přesnějších odhadů SW projektů.
- Výzkumné otázky na lineární závislost mezi odhady a skutečností.
- Stanovení parametrů UCP s největším vlivem na přesnost odhadu ve zhotoveném frameworku.
- Analýza vlivu účelových funkcí na přesnost odhadu.

Cíle byly rámcově splněny, jak to autor prokázal na experimentální části práce. Inovativním způsobem zkombinoval několik metod (zejména symbolická regrese a DE) do vlastní metody a dokázal je úspěšně aplikovat v praxi, což velmi oceňuji – práce má také významný dopad do praxe, nejde jen o akademický výstup. Taktéž analýza vlivu účelových funkcí může pomoci výzkumníkům z oblasti SW inženýrství pro použití efektivních výpočtů náročnosti projektů. To se týká i nalezených neefektivnějších kombinací parametrů UCP.

Přínos v oblasti poznání

Autorův přínos je hodnotný především z praktického pohledu vytvoření metod pro přesnější odhadování náročnosti projektů. Jeho inovativní spojení několika metod do hybridního frameworku má i svůj teoretický přínos, byť za hlavní výsledky považují prozkoumání vlastností parametrů, funkcí a jejich kombinací. I když byla metoda ověřena spíše na menším vzorku, je třeba zdůraznit, že k takovým datům je složité se dostat (firmy tyto zásadní manažerské údaje nechtějí volně šířit, což je pochopitelné).

Výsledky a další autorovy práce byly rovněž publikovány především na mezinárodních konferencích. Zde musím konstatovat jistý nedostatek v evidenci autorových publikací, u některých chybí základní údaje a není z nich tedy patrné, zda jde o sborník nebo časopis. To je dost podstatné k posouzení dopadu práce ve vědecké komunitě, proto bych doporučil, aby autor toto vysvětlil u obhajoby. Kvantitativní ukazatele publikační činnosti jsou přiměřené aplikačnímu charakteru práce. U některých publikací je vícero autorů, tudíž autor by měl u obhajoby jasně vymežit jeho přínos v těchto publikacích. Považuji to za velmi důležitou součást obhajoby. Dále bych také doporučil, aby autor označil několik nejvýznamnějších publikací vztahující se k tématu (část publikací se zřejmě netýká zaměření práce – např. Calibration of low-cost triaxial magnetometer).

Formální úprava, publikace

Doktorská disertační práce je členěna do 11 kapitol, včetně závěru a příloh. Další kapitoly pak obsahují literaturu a publikace autora. Po formální stránce je práce zpracována pečlivě, práce má jasnou strukturu a ilustruje výstižně použité metody i konkrétní vývoj. Obrázky výstižně popisují nejdůležitější poznatky získané experimenty i samotný Framework. Použitá literatura je přiměřená, obsahuje mnoho velmi aktuálních zdrojů (monografie, sborníky, časopisy).

Dotazy a připomínky

Celá práce je psána srozumitelně a postupně řeší problematiku od analýzy po konkrétní návrh. Autor by na obhajobě měl odpovědět na následující dotazy:

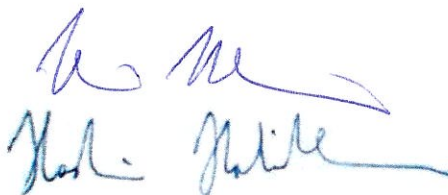
1. Jaký je podíl autora na uvedených publikacích se vztahem k tématu?
2. Které vlastní práce považujete za nejrelevantnější k tématu disertace?
3. Bylo by možné vygenerovat další datasey například simulací na uměle vytvořených datech. Případně jak budete postupovat dále s rozvojem metody, když je obtížné získat velké datasey pro experimenty.

Závěr

Předložená práce Ing. Tomáše Urbánka splňuje požadavky kladené na doktorskou disertační práci a to jak z pohledu teoreticko - metodologické úrovně, tak ve využitelnosti v praxi. Práce obsahuje původní výsledky.

Doporučuji předloženou disertační práci k obhajobě a rovněž doporučuji, aby na základě úspěšné obhajoby byl panu Ing. Tomáši Urbánkovi udělen akademický titul Ph.D.

V Ostravě 10. 11. 2020





Slezská univerzita v Opavě

Filozoficko-přírodovědecká fakulta v Opavě

Ústav informatiky

Bezručovo nám, 13, 746 01 Opava

Tel.: 553 684 368, fax: 553 684 370, e-mail: cs@fpf.slu.cz

Posudek disertační práce p. Ing. Tomáše Urbánka na téma „Využití analytického programování pro odhady časových náročností vývoje softwarových produktů“

Postup řešení

Předložená doktorská disertační práce je zaměřena do oblasti odhadu časové náročnosti vývoje softwarových projektů. Text práce je rozdělen do 9 kapitol, úvodu a závěru, celkem na 76 stranách. V kapitole 2 autor definuje hlavní cíl práce „Návrh frameworku k získání přesnějších odhadů časových náročností softwarových projektů“. Ke splnění definuje další podcíle spolu se stanovením výzkumných otázek. Výzkumné otázky se týkají lineární závislosti ohodnocení sw projektu a jeho skutečnou časovou náročností, vliv parametrů metody UCP (Use Case Points) na přesnější časový odhad, největší vliv různých fitness funkcí na přesnost odhadu, která fitness funkce poskytne nejpřesnější odhad. Další výzkumná otázka řeší porovnání odhadu standardní metodou UCP a navrženého frameworku. Poslední výzkumnou otázkou je fakticky porovnání navržené metody ve frameworku s 9ti různými metodami/algoritmy.

Kapitola 3 osahuje řešerši k řešené problematice. Kladně hodnotím závěr ohledně interpretovatelnosti analytického popisu oproti neuronovým sítím. V kapitole 4 autor zavádí metody použité v rámci disertační práce a v kapitole 5 popisuje použité metody vědecké práce.

Kapitolou 6 začíná jádro disertační práce, kde autor popisuje návrh frameworku, přičemž diagram navrhovaného frameworku, obr. 6.1. je principem návrhu. Nedílnou součástí je rovněž popis datových sad a jejich momentových, kvantilových, korelačních a dalších statistických charakteristik. Autor definuje podmínky experimentů pro jednotlivé výzkumné otázky. Agoritmické diagramy pro jednotlivé výzkumné otázky jsou vyjádřeny obrázky 6.4 – 6.8 a příslušným komentářem. Z mého pohledu je nejdůležitější omezení vygenerované rovnice na 13 elementů z důvodu komparace s lineární regresí. Autor rovněž uvádí, že by optimalizace s větším počtem členů mohla spadnout do lokálního extrému, což je pravda, nicméně v případě nalezení extrému globálního může být rovněž odhad přesnější. Je to vždy otázka kvality modelu vs. složitost struktury a pravděpodobnost nalezení globálního extrému



pro danou konfiguraci. Jeden z dotazů by mohl být, jaká bude situace v případě, že nebude stanoveno omezení počtu členů?

V kapitole 7 se musel autor poměřit s problémy statistické analýzy malého počtu pozorování, ale jedna realizace – sw projekt je otázka dlouhodobá 3-5 let, tedy není reálné získat stovky dat.

Posledním dotazem je otázka, zda do budoucna nezohlednit rovněž metodiku při vývoji projektu, tedy rigorozní vs. agilní, parametr typ metodiky vývoje sw?

Autor by mohl u srovnání jednotlivých optimalizačních metod s navrhovaným řešením detailněji popsat podmínky optimalizace. Příkladem může být, že u neuronových sítí není uvedena optimalizační metoda (gradient descent, levenberg *marquardt* aj.), jen základní algoritmus Backpropagation.

Závěrečné shodnocení:

Autor splnil cíle definované v kapitole 2:

- bod „2.1 Hlavní cíl“
- bod „2.2 Dílčí podpůrné cíle“
- bod „2.3 Stanovení výzkumných otázek“.

Diskutoval výsledky experimentů spolu s výzkumnými otázkami a navrhl další možné postupy zvyšování adekvátnosti navržené, implementované metody.

Po formální stránce je práce na vysoké úrovni, našel jsem jen drobnosti a to malý text u obrázků a v textu drobný překlep „Karnerovi metody“.

Z následujících důvodů

1. splnění cílů disertační práce,
2. jádro disertační práce přináší nové teoretické výsledky ve formě nové metody a algoritmu,
3. publikační výstupy vztažené k jádru disertační práce jsou na velmi dobré úrovni (v 9-ti publikacích jako hlavní autor),
4. formální zpracování na velmi úrovni,



Slezská univerzita v Opavě

Filozoficko-přírodovědecká fakulta v Opavě

Ústav informatiky

Bezručovo nám, 13, 746 01 Opava

Tel.: 553 684 368, fax: 553 684 370, e-mail: cs@fpf.slu.cz

doporučuji disertační práci pana Ing. Tomáše Urbánka na téma „Využití analytického programování pro odhady časových náročností vývoje softwarových produktů.“ k obhajobě v DRS oboru Inženýrská informatika. Po úspěšné obhajobě doporučuji udělit vědecký titul Ph.D.

Dne 19. 11. 2020 v Opavě.


Doc. Ing. Petr Čermák, Ph.D.

