

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Bc. Ksenia Shakurova
Oponent práce: doc. RNDr. Ing. Marcel Jiřina, Ph.D.
Název práce: Unsupervised Learning and Outlier Detection in Large Archives of Astronomical Spectra
Obor: Znalostní inženýrství

Datum vytvoření: 11. 6. 2016

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: Samotné zadání práce nevyznívá složitě, ale při praktické realizaci je třeba překonat řadu překážek, které jsou způsobeny rozsahem dat, časovou a paměťovou výpočetní náročností i obrovským rozsahem možností, co lze s daty dělat a jak volit postupy, metody a parametry.	
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 považuji za splněné. Ne všechny algoritmy byly implementovány pro paralelní výpočty, ale vzhledem k náročnosti zadání nevidím tento nedostatek za podstatný.	
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: Rozsah práce je cca 60 obsahových stran. Práce je informačně bohatá bez zbytečných částí.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	85 (B)
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: Práce má charakter průzkumného díla, kdy jsou diskutovány různé metody, ty jsou následně implementovány a aplikovány na reálná data. Díky složitosti úlohy, kdy z podstaty zadání existuje nepřeborné množství variant, co by s daty šlo dělat, je práce zvládnuta velmi dobře a podvýběr metod a zpracovávaných dat je volen velmi dobře. Zásadnější výhradu mám k použitému značení. V práci jsou pro různé pojmy (proměnné) využívány stejné symboly. Práce by měla mít jednotné značení a nemělo by docházet k "přetěžování" významu značení proměnných. V práci postrádám přehled metod pro detekci odlehlých hodnot. Uvedena je jen metoda LOF, která je v práci použita a okrajově je zmíněna metoda LOOP. Vzhledem k tomu, že studentka provedla v jiných částech práce velmi podrobnou rešerši různých jiných metod (např. metod pro shlukovou analýzu), vnímám toto jako nedostatek. Ve vztahu (1.1) na str. 9 by měl sčítací index "j" začínat od 1 a ne od 0.	

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
5. Formální úroveň práce	90 (A)
<i>Popis kritéria:</i> Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 12/2014, článek 3.	
<i>Komentář:</i> Práce je psána v angličtině a je zpracována na velmi dobré úrovni. Jazykovou stránku je pro mě jako nerodilého mluvčího obtížné posoudit, nicméně mám místy pocit, že v řadě případů není gramatická výstavba v pořádku.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
6. Práce se zdroji	90 (A)
<i>Popis kritéria:</i> 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 uvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.	
<i>Komentář:</i> Práce uvádí seznam 42 pramenů. U webových zdrojů chybí datum citace. V práci jsou odkazy na zdroje uváděny poměrně hojně a správně, nicméně na řadě míst, které uvádí některá tvrzení a pojmy, chybí citace zdroje.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
7. Hodnocení výsledků, publikační výstupy a ocenění	95 (A)
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Zvolené metody považuji za dobře vybrané. Implementace považuji za korektně zvládnuté. Ne všechny postupy byly díky omezeným možnostem použitých knihoven implementovány, což lze ale těžko vytýkat vzhledem k obrovskému záběru možností a z nich vyplývajících kombinací možných přístupů k řešení zadané práce. Práce skýtá potenciál pro publikovatelné vědecké výsledky.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - nehodnotí se</i>
8. Komentář o využitelnosti výsledků	
<i>Popis kritéria:</i> Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.	
<i>Komentář:</i> Zadání práce bylo formulováno s ohledem na konkrétní poptávku po zpracování astronomických dat a výsledky práce jsou v praxi využitelné.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - nehodnotí se</i>
9. Otázky k obhajobě	
<i>Popis kritéria:</i> Uvedte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).	
<i>Otázky:</i> Jaké existují přístupy a metody pro detekci odlehlých hodnot?	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
10. Celkové hodnocení	90 (A)
<i>Popis kritéria:</i> 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.	
<i>Text hodnocení:</i> Uvážíme-li šíři rozsahu zadání, je práce velmi dobře zvládnuta.	

Podpis oponenta práce: