FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Hodnocení vedoucího závěrečné práce

Student: Bc. Ondřej Podsztavek
Vedoucí: práce: RNDr. Petr Škoda, CSc.
Název práce: Neural Networks Based Domain Adaptation in Spectroscopic Sky Surveys
Obor: Znalostní inženýrství

Datum vytvoření: 27. 1. 2020

Hodnotící kritérium:	*Způsob hodnocení – následující škálou 1 až 4:*
1. 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 dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Komentář:

Zadání práce bylo splněno beze zbytku. Student zvládl všechny kroky zadání úspěšně a odhalil problém použití doménové adaptace na reálných datech z velkých astronomických archivů. Zadání práce je poměrně dosti náročné. Přímo navazuje na astroinformatickou problematiku řešenou na Astronomickém ústavu AVČR v Ondřejově. Student navazuje na svou bakalářskou práci zaměřenou na hledání zajímavých astronomických spekter emisních objektů v několika milionových přehlídkach projektu LAMOST a na publikaci zaslanou do impaktovaného časopisu Astronomy and Astrophysics, kde zkoumá hranice aktivního učení v dané problematice. V předkládané práci se zevrubně věnuje ověření použitelnosti moderní metodologie doménové adaptace na případ reálných astronomických spekter. Student se musel detailně seznámit s mnoha pojmy z fyziky, základy astronomické spektroskopie a s nástroji strojového učení, jako Tensorflow a Torch, stejně jako s mnoha knihovnami pro analýzu a preprocessing astronomických spekter ve formátu FITS, VO-Table a HDF5 (hlavně spojení s frameworkem Pandas).

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
2. Písemná část práce	*95 (A)*

Popis kritéria:
Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále 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. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Komentář:

Práce má 58 stran textu, 7 stran bibliografie, 10 stran příloh a 13 stran tiráže a rejstříků. Svým rozsahem tedy plně splňuje požadavky na magisterské práce.

Práce je logicky velmi dobře členěna do pěti kapitol. První tři shrnují kompaktní formou teoretické základy astronomické spektroskopie, spektroskopické přehlídky oblohy i principy a metody doménové adaptace, zejména v oblasti hlubokého učení. Pátá kapitola je shrnutí závěrů doporučení pro budoucí výzkum. Těžiště práce pak leží ve vlastní implementaci experimentů popsaných ve čtvrté kapitole. Je zde vše od představení testovacího vzorku, předzpracování dat, analýzu statistického rozložení dat v parametrickém prostoru metodami PCA, tSNE a UMAP až po vlastní experimenty se čtyřmi typy neuronových sítí navržených pro ověření jednotlivých reprezentativních metod. Výsledky jsou přehledně shrnuty v tabulkách záměn (confusion matrix).

Formální úroveň práce je velmi vysoká. Srozumitelný výklad je doprovázen názornými schématy a barevnými grafy. Všechny termíny a zkratky jsou řádně vysvětleny, rovnice korektně číslovány. Práce je psána celkem dobře srozumitelnou angličtinou s minimem překlepů, i když některé věty by před případnou publikací vyžadovaly lepší stylistiku a důslednou kontrolu gramatiky. Po typografické stránce lze vytknout zejména zbytečné odsazení textu po rovnicích.

Všechny použité zdroje jsou řádně citovány a myslím, že analýza dostupných pramenů je velmi podrobná Je citováno 62 pramenů, což svědčí o tom, že student provedl velmi rozsáhlý průzkum dostupné literatury i elektronických zdrojů. Citace jsou uváděny podle citačních standardů. Navíc jsou uvedeny i elektronické odkazy u některých hůře dostupných publikací. V souladu s moderními trendy jsou uváděny i DOI.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
3. Nepísemná část, přílohy	*95 (A)*

Popis kritéria:
Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů

Komentář:

Vedle tištěné části práce jsou velmi přínosné pro širší komunitu i Jupyter notebooky, které student připravil na přiloženém CD. Lze je výborně použít pro další analýzu nových dat. Zejména hotové modely neuronových sítí či procedury pro zpracovani spekter. Škoda je, že notebooky nejsou komentovány pro didaktické účely. Přesto jsou celkem pochopitelné.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Hodnocení výsledků, jejich využitelnost	*98 (A)*

Popis kritéria:
Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Komentář:

Výsledky práce mají velkou vědeckou hodnotu. Ukazují na správnost použití hlubokého učení ve velkých archivech astronomických spekter, což je dosud problematika velmi nová řešená i na školitelově pracovišti. I když se na první pohled zdá, že použité metody nefungují, je práce velmi cenná, neboť ukazuje na hlavní úskalí použití metod doménové adaptace na reálné datové archivy. Zde nejsou data idealně vyčištěna a zbavena artefaktů z důvodu jejich velkého objemu a nemožnosti vizuální kontroly všech. Cenné jsou ukázky špatně identifikovaných (falešně positivních zdrojových) spekter. Zde se při vizuální kontrole ukázalo, že i přes proklamovanou spolehlivou detailní analýzu kvasarů přehlídky SDSS se zde ukrývají dosud neznámé další kvasary. Naopak falešně pozitivní cílová spektra obsahují většinu LAMOST spekter poškozených, zdeformovaných i redukčních artefaktů.

Hodnotící kritérium:	*Způsob hodnocení – následující škálou 1 až 5:*
5. Aktivita a samostatnost studenta	5a: ***1=výborná aktivita,*** *2=velmi dobrá aktivita,* *3=průměrná aktivita,* *4=slabší, ale ještě dostatečná aktivita,* *5=nedostatečná aktivita* 5b: ***1=výborná samostatnost,*** *2=velmi dobrá samostatnost,* *3=průměrná samostatnost,* *4=slabší, ale ještě dostatečná samostatnost,* *5=nedostatečná samostatnost*

Popis kritéria:
V souvislosti s průběhem a výsledkem práce posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (5a). Posuďte schopnost studenta samostatné tvůrčí práce (5b).

Komentář:

Student pracoval velmi samostatně, dohledal si drtivou většinu použité literatury a dokumentace sám a zcela sám si i naprogramoval veškeré programy a skripty. Student se pravidelně účastnil všech konzultací (většinou jednou týdně) a reagoval operativně na e-mailovou korespondenci. Aktivně se učil i základy astrofyziky, zejména spektrální analýzy.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Celkové hodnocení	*96 (A)*

Popis kritéria:
Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

Text hodnocení:

Student vytvořil profesionálně vypadající práci s velkým publikačním potenciálem. Jako jeden z mála v oblasti astronomie se pokusil ověřit použitelnost doménové adaptace na reálné astronomické přehlídky spekter, které nelze člověkem klasifikovat kvůli jejich objemu. Práce je i pěkně napsána a doprovázena řadou praktických notebooků sloužících k řešení podobných problémů. Proto ji uděluji hodnocení A.

Podpis vedoucího práce: