



Posudek oponenta závěrečné práce

Oponent práce:	Ing. Daniel Vašata, Ph.D.
Student:	Nad'a Fučelová
Název práce:	Vizualizace chování algoritmů strojového učení po doplnění chybějících dat
Obor / specializace:	Umělá inteligence 2021
Vytvořeno dne:	10. června 2024

Hodnotící kritéria

1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Zadání bylo splněno bez výhrad.

2. Písemná část práce

85 / 100 (B)

Práce je logicky strukturovaná a po typografické a jazykové stránce velmi dobrá s pouze minimálním množstvím překlepů. Zdroje jsou relevantní. Co se týče způsobu citování, nelíbí se mi způsob, kdy se citace uvádí až na konci věty s tím, že někdy je to za tečkou a někdy před tečkou. Hlavní výtky k textu mám však po obsahové stránce. Terminologicky se mi nelíbí používání pojmu chybějící datový bod. Termínem chybějící datový bod by se neměl označovat případ, kdy u nějakého datového bodu (vektoru v prostoru příznaků) chybí hodnota nějakého z příznaků. Ve vzorci (1.1) je použit symbol psí, který není nikde definován. Dále je v tomto vzorci citace, což je velmi nestandardní. V poslední větě v části 1.1 se píše, že "Odhadované parametre nie sú nijak ovplyvnené absenciou údajov.", což není jasné, jak je myšleno, neboť není nikde definováno, co jsou "odhadované parametry". Na konci sekce 4.2, je špatně uvedeno, že článek [24] se zabývá absencí celých příznaků. V takovém případě by samozřejmě nebylo možné žádná data doplňovat, neboť by nebylo odkud. Vztah (5.1) není Eukleidovská vzdálenost. V sekci 5.2 je uvnitř zmíněno, jak se postupuje v případě kategorických příznaků. V takovém případě ale není vhodné metodu nazývat imputace průměrem. V části 5.4 je nesrozumitelně popsáno, co je v proměnných y_{obs} , y_{miss} , x_{obs} a x_{miss} , ani jakého jsou přesně typu (vektory, matice apod.) a zda se to bere pro každý datový bod zvlášť, nebo pro celý dataset.

Bez ohledu na výše uvedené výtky je však text práce na velmi slušné úrovni.

3. Nepísemná část, přílohy

100 /100 (A)

Nepísemnou částí práce je webová aplikace pro analýzu a vizualizaci chybějících dat a metod jejich doplňování. Vytvořená aplikace je pěkná a plně funkční. Zdrojové kódy jsou srozumitelné a dobře komentované.

4. Hodnocení výsledků, jejich využitelnost

95 /100 (A)

Teoretickým výsledkem práce je popis problematiky doplňování chybějících hodnot, který má některé výše uvedené problémy, Celkově je ale na velmi slušné úrovni a může přinést zájemcům o danou problematiku základní vhled. Praktickým výsledkem práce je pak vytvořená webová aplikace, která je velmi pěkná. Může sloužit zejména pro výukové účely, ale dovedu si představit její užití i v některých komerčních aplikacích.

Celkové hodnocení

90 /100 (A)

Práce je celkově na velmi slušné úrovni. Vzhledem k využitelnosti hlavního výsledku tak i přes některé nedostatky v teoretické části navrhuji hodnocení stupněm A.

Otázky k obhajobě

Co znamená symbol psi ve vzorcích (1.1)-(1.3) a v obrázku 1.1?

Umí Vaše aplikace pracovat s různými reprezentacemi chybějících hodnot? Například, pokud bude chybějící hodnota jinak číselného příznaku v csv souboru označena textem "NaN".

Instrukce

Splnění zadání

Posudte, 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, posudte 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.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, 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. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, 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.

Nepísemná část, přílohy

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.

Hodnocení výsledků, jejich využitelnost

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.

Celkové hodnocení

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.