



Posudek oponenta závěrečné práce

Oponent práce: Ing. Karel Klouda, Ph.D.
Student: Olena Marchenko
Název práce: Kontextuální pasivní DNS prediktor
Obor / specializace: Znalostní inženýrství
Vytvořeno dne: 19. srpna 2021

Hodnotící kritéria

1. Splnění zadání

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

Všechny body zadání byly nějakým způsobem splněny, i když např. u bodu 3 trochu jinak, než je v zadání popsáno (data byla získána od společnosti Cisco). Zadání bylo poměrně pracné, komplikované i velikostí zpracovávaných dat. Z pohledu strojového učení šlo ale o poměrně jasně definovanou úlohu s automaticky získanými daty, která nevyžadovala složité zpracovávání, čištění a prozkoumávání.

2. Písemná část práce

65 /100 (D)

Práce je psána obstojnou angličtinou, ale kvalita textu je hodně nevyrovnaná. Některé pasáže jsou povedené, např. popis problému v první kapitole je dobře pochopitelný a neobsahuje mnoho chyb a nedostatků. Naopak popis experimentu v kapitole 3 je často nesrozumitelný a obsahuje chyb a nesrovnalostí hodně. Kapitola 2, která obsahuje víceméně zrychlený úvod do strojového učení, je z mého pohledu nadbytečná a formálně ne zcela zvládnutá.

Zopakovat na pár stránkách úvod do strojového učení vč. zavedení jednotného značení je těžký úkol a je otázka, jestli má smysl. Navíc zdaleka ne vše, co je vyloženo, je potřeba v dalším textu. Způsob výkladu, kdy je složité metodě věnován odstavec či dva, má dost sporný přínos: Kdo metodě rozumí, nepotřebuje jej, kdo nikoli, tomu takový výklad nepomůže. Celkově bych preferoval na úkor druhé kapitoly rozšířit kapitolu 3 s prezentací výsledků, doplnit ji příklady a přesnějším popisem nastavení experimentu.

V práci je též mnoho formálních chyb. Např. malá písmena na začátku vět (na několika místech), zvláštní členění do odstavců (např. začátek kapitoly 3 na str. 31) divné reference na obrázky a jiné entity (např. reference 3.3 na str. 35, nefunkční citace [?] na str. 20) a

mnoho nedorazů v matematickém značení a formálním popisu vč. referencí. Např. na str. 36 není jasné, co za objekty je v množině A. Ve vzorci dole na téže stránce je pak první h s indexy nahoře i dole, poslední už má ale index jen dole. Takových věcí je v práci hodně. Popisky grafů také nejsou vždy srozumitelné. Např. v mnoha grafech je na ose x "Classifier index", ale není nikde vyjasněno, co to znamená.

3. Nepísemná část, přílohy

75/100 (C)

Nepísemná část práce je sada skriptů zejm. v jazyce Python, která využívá standardní knihovny pro strojové učení. Kód je docela přehledně organizován, vezmeme-li v úvahu, že jde o experimenty. Mnoho metod a funkcí není doplněno komentářem, nebo se komentář sestává pouze z řetězce "Docstring".

4. Hodnocení výsledků, jejich využitelnost

73/100 (C)

Využitelnost výsledků se hodnotí těžko, protože jednou z největších slabín práce je nejasný popis experimentu a vyhodnocování výsledků. Zdá se, že použité metody dosahují lepších výsledků, než prosté odhady založené na frekvencích domén.

Celkové hodnocení

70/100 (C)

Autorka odvedla zřejmě hodně práce, ovšem kvalitu výsledků snižuje horší a nejasná prezentace. Navrhují proto hodnotit práci známkou C.

Otázky k obhajobě

1. Ve své prezentaci podrobně popište, jak přesně funguje Váš model. Proč jste si zvolila tento způsob, kdy vybíráte z konečného seznamu (např. proč je target variable u hostname prediction k+1 členný vektor a jak bylo voleno k)?
2. Na straně 32 píšete, že feature extraction bylo děláno pro trénovací a testovací množinu nezávisle. Můžete toto vysvětlit na příkladu příznaku is_in_top, jak přesně byl tento příznak konstruován?

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.