

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Bc. Ondřej Nový
Oponent práce: Ing. Daniel Vašata, Ph.D.
Název práce: Doporučovací systém pro výběr volitelných předmětů
Obor: Znalostní inženýrství

Datum vytvoření: 5. 6. 2017

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: Student se měl v rámci práce seznámit s metodami doporučování volitelných předmětů a navrhnout a posléze implementovat několik vhodných modelů. Na závěr měl vyhodnotit jejich přesnost, vybrat nejvhodnější a provést uživatelské testování tohoto modelu.	
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: Po rešeršní části práce se student zaměřil na dvě metody: kolaborativní filtrování založené na matici a kolaborativní filtrování založené na dopředné neuronové síti. Ačkoliv se z výsledků práce zdá, že metoda založená na neuronové síti dává z pohledu doporučení předmětů lepší výsledky (cca 2x lepší), autor práce k uživatelskému testování vybral metodu založenou na matici. Důvodem byla jednodušší interpretace doporučení studenty, protože metoda založená na matici zároveň predikuje známku, kterou by student z daného předmětu mohl obdržet. Tento přístup považuji za mírný nedostatek práce a podrobněji se k němu vyjadřuji v bodě 7. tohoto posudku.	
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: Zcela v pořádku.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	75 (C)
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 přehledná a srozumitelná. Větším nedostatkem je absence teoretických popisů použitých metod a jejich parametrů. Většina práce je věnována popisům problematiky doporučování, popisů zdrojových dat a jejich předzpracování a popisu výsledků. Samotné metody však nejsou nikde podrobněji definovány. V části 1.5.2. je uveden pouze obecný princip kolaborativního filtrování. Použitý algoritmus kolaborativního filtrování založeného na neuronové síti však již popsán není. Ani není nijak odkazován. Dále nejsou uvedeny definice různých hyperparametrů modelu. Např. není definován pojem kosinové vzdálenosti a dalších ztrátových funkcí. V části 6.7.5 je sice u ztrátových funkcí uveden odkaz, je zde ovšem odkazován pouze zdrojový kód použité knihovny v jazyce Python. Ještě horší je to u dalších hyperparametrů jako aktivační funkce a optimalizátor. Zde již nejsou uvedeny ani žádné odkazy. Na závěr mám také výtka ke grafům 6.2, 6.3 a 6.5 a 6.6, kde by bylo správné normovat jejich plochu na 1 (podobně jakou u hustoty pravděpodobnosti), aby bylo možné je porovnávat. Tak jak jsou udělány nyní to příliš nelze.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

5. Formální úroveň práce

90 (A)

Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 14/2015, článek 3.

Komentář:

Práce je jazykově a formálně na velmi slušné úrovni. Jediným systematickým nedostatkem je psaní tvrdého y ve spojení "pro budoucí profesey".

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

6. Práce se zdroji

79 (C)

Popis kritéria:

Vyjáďřete 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 úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Práce se zdroji je na standardní úrovni. Místo zdrojů [22, 23, 29] by bylo vhodnější zvolit knižní případně časopisecké publikaci vědeckého charakteru. Také by bylo vhodné doplnit relevantní citace na definice použitých modelů a jejich hyperparametrů.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

7. Hodnocení výsledků, publikační výstupy a ocenění

75 (C)

Popis kritéria:

Vyjáďřete 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.

Komentář:

Hlavními výsledky jsou: rozsáhlá analýza problematiky doporučování volitelných předmětů, struktury dostupných dat a provedené uživatelské testování vybrané doporučovací metody. Co se mi ale osobně nelíbí je logika, jakou byl vybrán finální model pro doporučování. Ačkoliv se z částí 6.7.6 a 6.6.3 zdá, že model založený na neuronové síti je skoro 2x lepší než model založený na matici, student se rozhodl vybrat jako finální model založený na matici. Hlavním argumentem je, že "model založený na matici mnohem lépe odhaduje známky z předmětů, což lze považovat za důležitý aspekt pro to, aby studenti chápali, jak doporučování funguje." Cílem práce je podle zadání samotné doporučování předmětů a nikoli doporučování známek. Je tedy otázka, zda je správné vyřadit model, který je z pohledu hlavního cíle (doporučování předmětů) mnohem lepší. Navíc si nejsem jist, zda z predikce známek studenti opravdu pochopí, jak doporučování funguje. Pochopení je podle mě spjato s porozuměním algoritmu doporučování a ne jeho výstupům (tj. predikcím známek).

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.

Komentář:

Výsledky jsou potenciálně využitelné pro studenty FIT a mohou jim sloužit k výběru volitelných předmětů.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

9. Otázky k obhajobě

Popis kritéria:

Uvedte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

Proč jste do části Ohodnocení nezahrnul i model založený na neuronové síti? Pokud byste tam totiž zahrnul oba modely, mohl byste provést finální volbu na základě významnější zpětné vazby a testování.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

10. Celkové hodnocení

80 (B)

Popis kritéria:

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.

Text hodnocení:

Předkládaná diplomová práce je lehce nadprůměrná a proto navrhuji hodnocení stupněm B.

Podpis oponenta práce: