



Posudek oponenta závěrečné práce

Student: Bc. Andrej Oliver Chudý
Oponent práce: doc. RNDr. Pavel Surynek, Ph.D.
Název práce: Detekce podobností datových domén pomocí metod strojového učení
Obor: Znalostní inženýrství

Datum vytvoření: 3. 6. 2019

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Cílem bylo vytvořit systém pro efektivní vyhledávání podobných sloupců v rozsáhlých databázích a provést jeho experimentální analýzu. Zadání nespecifikuje žádnou konkrétní oblast technik, kterou by měl uchazeč využít. Uchazeč zvolil použití rekurentních neuronových sítí ke kompresi vektoru příznaků, což vedlo k lepším výsledkům, než dosahuje konkurenční metoda TF-IDF. Navíc byl uchazečův návrh transferován do komerční aplikace Ataccama One. Zadání považuji za splněné.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
2. Písemná část práce	85 (B)
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Práce ihned přechází k návrhu řešení pomocí hlubokého učení a rekurentních neuronových sítí, aniž by byly důkladně diskutovány alternativy, které například vůbec neuronové sítě nepoužívají. Nabízí se třeba zvážení algebraických metod, tj. maticové komprese příznakových vektorů atd. Kapitola 2, která názvem budí dojem, že v ní tato analýza vhodnosti zvoleného řešení vzhledem k alternativám bude podána, toto vůbec nečiní a je celkově slabší. Další kapitoly věnující se vlastnímu návrhu mají lepší úroveň, čekal bych zde četnější odkazy na použitou literaturu, Jsem spokojen s experimentální částí práce (kapitoly 5, 6). Alternativní metoda TF-IDF je zde zmíněna, podrobná teoretická konfrontace TF-IDF vůči navržené metodě s neuronovými sítěmi mi ale chybí v předchozích kapitolách. Vlastní výsledky jsou přesvědčivé. Vůči existujícím metodám bylo dosaženo zlepšení asi 14%. Prezentaci vlastních výsledků by šlo lépe oddělit od existujících metod, aby přínos lépe vynikl.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
3. Nepísemná část, přílohy	95 (A)
<i>Popis kritéria:</i> 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, Výzumná a experimentální práce – opakovatelnost experimentů	
<i>Komentář:</i> Implementace metod odpovídá popisu v textové části. Metody fungují a celkově jsem s nepísemnou částí spokojen.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>

4. Hodnocení výsledků, jejich využitelnost

95 (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:

Uchazeč výsledky své práce transferoval do praxe a nyní se používají v komerčním softwaru Ataccama One, což hodnotím kladně.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

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

Otázky:

Uchazeč by mohl pojednat o alternativních metodách, které nepoužívají neuronové sítě, a proč zvolil přístup založený na neuronových sítích.

Hodnotící kritérium:

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

6. Celkové hodnocení

85 (B)

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:

Práce je kvalitní, nedostatkem je pouze slabší teoretická analýza vzhledem k existujícím metodám, silnou stránkou je naopak praktická orientace práce a uskutečněný transfer výsledků do praxe. Práci doporučuji přijmout k obhajobě jako magisterskou. Znamku zatím navrhuji B s tím, že pokud se uchazeč u obhajoby uspokojivě vypořádá s připomínkami, existuje vzhledem k potenciálu práce prostor pro zlepšení hodnocení.

Podpis oponenta práce: