

HODNOCENÍ BAKALÁŘSKÉ PRÁCE

Oponent BP

Jméno bakaláře: Robert Brada

Garantující katedra: KKY

Název bakalářské práce: Identifikace rodného jazyka pisatelů na základě anglicky psaných esejů

	Předmět hodnocení	Nadprůměrné	Průměrné	Podprůměrné
1	Jazyková a grafická úprava	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Formální a obsahová stránka práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3	Vhodnost použitých metod	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Způsob zpracování a vyhodnocení	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Správnost získaných výsledků	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	Vlastní přínos	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
7		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Doplnění hodnocení, připomínky, dotazy:

Autor práce se věnuje úloze identifikace rodného jazyka z textů psaných v jiném jazyce. Pro tuto identifikaci používá standardních metod obvyklých v oboru, jedná se především o TF-IDF příznaky a support vector machines. V práci uvádí i teoretický popis těchto metod, nicméně tento popis je z mého pohledu pouze částečný - chybí popis zobecnění SVM klasifikátorů na lineárně neseperabilní problémy a rozšíření pro klasifikaci do více tříd. Tento nedostatek se projeví v experimentální části práce, kde autor operuje s regularizačním parametrem C, který je zaveden právě při výše zmíněném zobecnění.

Jako pozitivní hodnotím snahu autora o detailní analýzu získaných výsledků, ať již se jedná o vliv příznaků na klasifikaci nebo křivky učení. Zde používá metodu LIME (Local Interpretable Model-agnostic Explanations), bohužel neuvádí detailnější popis metody ani citaci relevantní publikace.

I přes zmíněné nedostatky práce více než dostatečně splňuje všechny body zadání, práci doporučuji k obhajobě s hodnocením velmi dobře. Do diskuze k obhajobě práce doporučuji zařadit následující body, které více objasní některé pasáže práce:

- Na straně 28 je zmíněno spojení trénovací a testovací sady za účelem trénování doc2vec modelu. Je toto spojení korektní? Jak by se postupovalo, pokud by klasifikátor měl zpracovat nová, dosud neviděná data?

- Optimalizaci metaparametrů provádíte pro metaparametry loss a C (programový kód na straně 30). Jaký je význam těchto metaparametrů? Jak jste dospěl k množinám hodnot, z nichž následně vybíráte optimální hodnoty? Jak následně probíhá výběr optimálních parametrů?

Splnění bodů zadání úplně částečně nesplněno

Doporučení práce k obhajobě ano ne

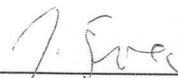
Celkové hodnocení práce výborně velmi dobře dobře nevyhověl

Jméno, příjmení, titul oponenta BP: Ing. Jan Švec, Ph. D.

Pracoviště oponenta BP: KKY

14.6.2018

Datum


Podpis