

Jméno bakaláře: Jan Kandyba

Garantující katedra: KKY

Název bakalářské práce: Detekce tématu dokumentu

	Předmět hodnocení	Nadprůměrné	Průměrné	Podprůměrné
1	Jazyková a grafická úprava	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	Formální a obsahová stránka práce	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3	Vhodnost použitých metod	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Způsob zpracování a vyhodnocení	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	Správnost získaných výsledků	<input checked="" type="checkbox"/>	<input 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 se ve své bakalářské práci věnuje detekci tématu z textových dokumentů. V teoretické části popisuje použité metody a v experimentální části je pak vhodně vyhodnocuje na reálných datech. Autor správně natrénoval vybrané klasifikátory textu a vygeneroval množství experimentálních výsledků, které jsou nezpochybnitelným přínosem pro aktuálně řešený projekt NAKI.

Nedostatky práce spatřuji zejména po formální a obsahové stránce. Teoretický popis použitých klasifikátorů je často nepřesný. Autor operuje s pojmy, které jsou definovány až dále v textu nebo nejsou definovány vůbec (např. parametrizace, minimální a maximální DF). Orientaci v dosažených výsledcích komplikuje chybné číslování tabulek a obrázků ke konci práce. V práci postrádám také bližší popis použitých dat a témat. Přiložený obrázek je špatně čitelný a k lepšímu pochopení struktury dat příliš nepřispívá.

I přes uvedené nedostatky autor jednoznačně prokázal schopnost samostatně nastudovat moderní metody strojového učení a aplikovat je na reálná data. Práci proto hodnotím jako velmi dobrou a doporučuji ji k obhajobě.

Otázky:

1. Výsledky pro různé hodnoty dropout (tab. 10.18) jsou interpretovány tak, že nejvyšší přesnosti klasifikace je dosaženo, když se při trénování „90% vstupních dat náhodně vymaže“. Můžete tento fenomén podrobněji vysvětlit?
2. Problém přetrénování neuronových sítí byl v práci vyřešen experimentálním nalezením optimálního počtu trénovacích epoch (120). Jaké další metody lze použít proti přetrénování?
3. Práce pracuje s předpokladem, že třídy v datech musí být vyvážené, čehož bylo dosaženo odebráním dat z frekventovaných tříd. V mnoha úlohách ale může být zahazování anotovaných dat nežádoucí. Jak by se dalo trénování a vyhodnocení upravit, aby bylo možné použít kompletní nevyvážená data?

Splnění bodů zadání	<input checked="" type="checkbox"/> úplně	<input type="checkbox"/> částečně	<input type="checkbox"/> nesplněno
Doporučení práce k obhajobě	<input checked="" type="checkbox"/> ano		<input type="checkbox"/> ne
Celkové hodnocení práce	<input type="checkbox"/> výborně	<input checked="" type="checkbox"/> velmi dobře	<input type="checkbox"/> dobře <input type="checkbox"/> nevyhověl
Jméno, příjmení, titul oponenta BP: Ing. Jan Lehečka			
Pracoviště oponenta BP: NTIS			

10. 6. 2019

Datum