

Posudek oponenta bakalářské práce

Autor práce: **Tereza KAŠÁKOVÁ**

Název práce: **Automatické rozpoznávání typů otázek v různých jazycích**

Jazyková a grafická úprava

Průměrné

Formální a obsahová stránka práce

Průměrné

Vhodnost použitých metod

Nadprůměrné

Způsob zpracování a vyhodnocení

Nadprůměrné

Správnost získaných výsledků

Nadprůměrné

Vlastní přínos

Nadprůměrné

Doplnění hodnocení, připomínky:

Bakalářská práce se věnuje rozpoznávání typu otázek v různých jazycích, a to z jejich textové podoby. Konkrétně se autorka zabývala klasifikací otázek na otázky se stoupavou nebo klesavou intonací, což má posloužit ke zlepšení počítačové syntézy řeči.

Teoretická část práce je psána adekvátně, jak z obsahového tak z formálního hlediska. Drobné výhrady bych měla pouze k seznamu referencí, kde mi několik z nich nepříjde pro daný kontext zcela ideální (např. na str. 3, v kontextu využití neuronových sítí pro syntézu řeči, by bylo vhodnější místo ref. [9] z roku 1996 citovat něco aktuálnějšího). Celkově ale reference odpovídají očekávané úrovni u bakalářských prací.

Hlavním přínosem bakalářské práce pak je aplikace několika existujících metod strojového učení na zadanou úlohu. Autorka využila několik rozšířených metod, které jsou k tomuto účelu vhodné, od prostého pravidlového přístupu až po velmi nový model neuronových sítí T5. V případě pravidlového přístupu autorka navíc sama navrhla dle mého názoru velmi rozumně znějící pravidla a to jak ručně, tak na základě automatické analýzy dat. Také se mi líbí, že v závěru práce je analýza nejčastějších chyb a několik návrhů jak by výsledky bylo možné dále zlepšit.

Co jsem bohužel v práci nenalezla je zmínka odkud použitá data pocházela - pokud se například jedná o nějaké veřejně dostupné datasety, bylo by vhodné toto uvést. Velmi se mi naopak líbí, že zdrojový kód, data i natrénované modely jsou veřejně přístupné na autorčině GitHubu.

Z hlediska samotného textu bych v experimentální části vytkla pouze to, že v sekci 5.4 ("Neuronová síť T5") není vždy na první pohled zřejmé kdy autorka popisuje vlastní práci a kdy práci původních autorů sítí T5. Objasňuje to až poznámka pod čarou na konci sekce. Ve shrnutí výsledků (kapitola 6, strana 39) se také objevuje pojem "Tf-idf transformace", který ale jinde v práci nebyl vysvětlen (objevuje se pouze zmínka o "instanci TfIdfTransformer()") - tohoto se také týká má otázka.

Zmíněné nedostatky ale nijak významně neubírají z celkově velmi pozitivního dojmu který práce zanechává. Hodnotím ji tedy výborně.

Dotazy

1. Ve shrnutí výsledků se objevuje pojem "Tf-idf transformace", který ale jinde v práci nebyl zmíněn. Můžete lépe objasnit co zkratka "Tf-idf" znamená a o co se jedná?
2. Uvádíte, že pro ruská data nebyl ručně vytvořen seznam tázacích zájmen a proto pro ruský jazyk nejsou k dispozici výsledky s ručně odvozenými pravidly. Zkoušela jste tento seznam vytvořit automatickým překladem z jiného jazyka?

Splnění bodů zadání

úplně

Doporučení k obhajobě

ANO

Hodnocení: 1 - Výborně

V _____ dne _____

Ing. Marie Kunešová, Ph.D.