
Posudek oponenta bakalářské práce

Jakub Mladý
Metody Active learning pro úlohy automatického zpracování přirozeného
jazyka

Cílem práce Jakuba Mladého bylo prozkoumat metody aktivního učení (active learning) vhodné pro úlohy v oblasti zpracování přirozeného jazyka a navrhnout sadu experimentů pro jejich vyhodnocení na několika datových sadách.

Práce má 33 stran původního textu. Obsahuje velmi malé množství gramatických chyb a překlepů. Práce se však poměrně obtížně čte protože obsahuje složité větné konstrukce a chyby ve stavbě vět. Autor také používá překlady ustálených anglických termínů bez uvedení originálu tam, kde neexistuje zažitý český překlad. Gramatická úroveň práce je nevyvážená. V závěru práce se vyskytuje daleko více chyb než v jejím zbytku. Práce je vhodně doplněna precizně vysázenými vzorci. Ocenil bych kdyby v práci bylo více obrázků a schémat tam, kde je to vhodné. Z typografického hlediska nemám k textu žádné zásadní výhrady. Jen grafy s výsledky jsou tak malé, že jsou v tištěné formě prakticky nečitelné.

Ke struktuře a obsahu práce mám také několik výhrad. Členění textu do kapitol není zvoleno vhodné; V teoretické části mi chybí kapitola zabývající se strojovým učením a neuronovými sítěmi. Použité modely jsou pak velice stručně vysvětlovány v kapitole s popisem experimentů. Na druhou stranu, kapitola 3, která se zabývá teorií aktivního učení (stěžejní část teorie) je zpracovaná velice kvalitně.

K experimentální části mám také několik výhrad, zejména k prezentaci výsledků. V textu mi chybí statistiky použitých datových sad o které by se mohl poté autor opřít v detailnější diskuzi nad výsledky. Také mi chybí přímé srovnání baseline modelů (bez aktivního učení) s publikovanou literaturou (pro kontrolu, že není chyba v autorově implementaci). V diskuzi mi chybí odkazy na konkrétní grafy které podporují prezentované závěry. Pozitivně hodnotím množství provedených experimentů a jejich realizaci v prostředí metacentra.

Student ve své práci cituje dostatečné množství relevantních zdrojů, tištěných i webových. Jeho práci z literaturou nemám co vytknout.

Vytvořený programový kód je plně funkční, srozumitelný a vše je detailně popsáno v příložených readme souborech. Z příložených souborů je patrné, že autor odvedl velké množství práce

Zadání bylo splněno ve všech bodech, ačkoliv některým bodům se měl autor věnovat o něco více (zejména zhodnocení dosažených výsledků, viz připomínky výše). Přes výše zmíněné připomínky považuji práci jako celek za dobrou.

Práci proto **doporučuji** k obhajobě hodnotím klasifikačním stupněm

„dobře“.

Doplňující otázky:

1. Z grafů je vidět, že pro některé datasety modely je výrazně větší rozptyl úspěšnosti než pro jiné. Čím je to podle Vás ovlivněno?
2. Jaký je přínos aktivního učení v případě, že použijete pouze velmi malou část datové sady (desítky až stovky vzorků)?

Ing. Ondřej Pražák
(oponent BP)

V Plzni 20. srpna 2023