

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Patrik Romanský
Název práce Automatická detekcia fake-news na slovenských textoch
Rok odevzdání 2023
Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Michal Novák **Role** oponent
Pracoviště ÚFAL MFF UK

Text posudku:

Popis práce: Predložená diplomová práce sa venuje úlohe detekcie tzv. fake-news, čiže nepravdivých článkov v slovenských žurnalistických textoch (pre zjednodušenie bez ohľadu na úmysel).

Za týmto účelom autor práce využíva predovšetkým pracovnú verziu verejnej dátovej sady SlovakCovid19 FN od Sarnovského a kol. (2022), ktorá obsahuje články súvisiace s ochorením Covid-19. Okrem toho pracuje ešte s jeho podmnožinou, ktorú sám vybral tak, aby bol podiel pravdivých a nepravdivých článkov vyvážený. Pri niektorých experimentoch autor využíva aj dataseť na rozpoznávanie fake-news v angličtine, konkrétne korpus Covid19 FN s rovnakou tematikou článkov ako v prípade slovenských dát a korpus LIAR, ktorý obsahuje najmä články týkajúce sa volieb v USA.

Pri detekcii autor experimentuje s širokým spektrom algoritmov strojového učenia: od klasických (napr. rozhodovacie stromy, Naive Bayes) cez neurónové siete (napr. konvolučné, rekurentné) až po adaptáciu predtrénovaných veľkých jazykových modelov (napr. BERT, RoBERTa, XLM-RoBERTa, SlovakBERT), tzv. fine-tuning. Autor navrhuje a testuje niekoľko metód fine-tuningu. Tie kombinujú mono- i multilinguálne jazykové modely s vyššie spomenutými korpusmi v slovenčine a angličtine pokrývajúcimi dve odlišné témy (Covid-19 a voľby v USA), pričom v prípade potreby rozlišuje, či pri adaptácii použil autentické dáta alebo ich preklady prostredníctvom služby Google Translate.

Práca je písaná po slovensky. Má 144 strán vrátane textovej prílohy, zoznamu použitej literatúry, obrázkov a tabuliek. Súčasťou práce je aj dátová príloha, ktorá obsahuje zdrojové kódy experimentov a použité dataseť.

Hodnotenie: Táto práca je druhou odovzdanou verziou, pričom tú prvú som k obhajobe nedoporučil. Z toho dôvodu vo svojom posudku budem tieto dve verzie porovnávať, prípadne sa odkazovať k aspektom, ktoré som práci vyčítal v prvom posudku.

Na začiatok by som vyzdvihol silné stránky práce. Jednou z nich je nemalé úsilie vložené do čistenia a vyváženia slovenského korpusu SlovakCovid19 FN. Autor odstránil z pôvodného korpusu tie časti, ktoré boli chybné zahrnuté do obsahovej zložky článkov, pričom sa jedná skôr o metadáta k článkom alebo tzv. boilerplate (napr. meno autora, jednotné hlavičky a pätičky identifikujúce zdroj článkov). Prítomnosť týchto prvkov v texte úlohu automatického rozpoznávania falošných správ príliš uľahčuje, na čo aj autor správne upozorňuje pri porovnaní svojho prístupu s experimentami uskutočnenými samotnými autormi korpusu. Po mojej kritike predchádzajúcej verzie práce je táto časť aj dostatočne jasne vysvetlená a zdôraznená.

Druhou silnou stránkou je pestrá sada uskutočnených experimentov, pričom autor postupoval tak, ako sa na prvú rozsiahlejšiu štúdiu v tejto oblasti patrí. Začal od klasických metód strojového učenia v oblasti spracovania prirodzeného jazyka a postupne dospel až k najmodernejším metódam založených na veľkých jazykových modeloch. Do kľúčovej kapitoly venujúcej sa experimentom s jazykovými modelmi a ich adaptáciou autor pridal analýzu výkonnosti týchto modelov z troch rôznych pohľadov: jazyk, doména a zložitosť modelu. Tým zhŕňa pozorovania zo širokej palety experimentov do pomerne jednoduchých odkazov, ktoré si čitateľ z tejto práce môže vziať. Hoci by som si vedel v analytickej časti predstaviť ďalšie podporné analýzy alebo experimenty, aj v aktuálnej podobe je táto časť významným príspevkom k celkovej koherencii odovzdanej práce.

Veľmi oceňujem pridanie kapitoly 14 popisujúcej obmedzenia práce. Autor sa tak nebojí okrem vyzdvihnutia silných stránok svojej práce priznať tie slabšie.

Tým sa dostávam k prezentácii výsledkov práce, ktorá bola z môjho pohľadu v prvej verzii diplomovej práce nevyhovujúca po viacerých stránkach. Musím uznať, že na prezentácii autor zapracoval a väčšinu mnou kritizovaných chýb a nedostatkov opravil. Dôležité detaily týkajúce sa režimu evaluácie pri rôznych experimentoch sú už teraz jasné. Na rozdiel od prvej verzie su aj formálne chyby v texte, grafoch a tabuľkách výrazne eliminované do miery akceptovateľnej pre diplomové práce.

Z môjho pohľadu najväčším nedostatkom stále ostáva na diplomové práce neobvykle mohutný rozsah práce, ktorý sa oproti prvej verzii ešte zväčšil. Ten je spôsobený širokým spektrom odskúšaných metód strojového učenia (čo by som kritizovať skôr nemal), ale aj neúsporným formulovaním myšlienok. Hoci mnohé opakovania a nadbytočné pasáže z prvej verzie práce autor odstránil, dalo by sa v tomto smere zájsť určite ďalej.

Na záver ešte pridám niekoľko konkrétnych komentárov:

1. Metóda XGBoost na rozdiel od predchádzajúcej verzie je v tej aktuálnej aspoň popísaná (v sekcii 6.4.2). Namiesto “marketingového” popisu vlastností vylepšenej verzie algoritmu Gradient Boosting by som sa však radšej dozvedel viac o princípe jeho fungovania.
2. V sekcii 13.3.1 sa nedostatočné výsledky (zrejme všetky články označené ako pravdivé/nepravdivé)

multilinguálnych modelov na nepreložených dátach s postupným fine-tuningom na anglickej dátovej sade LIAR a slovenskej sade SlovakCovid19 FN pripisujú striedaniu jazykov. Pritom po nahradení sady LIAR inou anglickou sadou COVID19 FN výsledky vyzerajú lepšie.

3. Pri analýze vplyvu komplexity modelov v sekcii 13.7 mi chýba zmienka o tom, že SlovakBERT nevdáčí svojej najlepšej výkonnosti len samotnej architektúre Transformer, ale aj tomu, že je na rozdiel od modelov LSTM a XGBoost predtrénovaný na obrovskom množstve neanotovaných textov.

Záver: Riešiteľ svojou diplomovou prácou ukázal, že je do uspokojivej miery schopný metodologicky správne navrhnuť a uskutočniť experimenty v oblasti strojového učenia a spracovania prirodzeného jazyka. Po odstránení väčšiny chýb a nedostatkov v prezentácii z prvej verzie sa aj tento aspekt práce zlepšil. Predloženú diplomovú prácu preto doporučujem k obhajobe.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 28. 8. 2023

Podpis: