

Posudek oponenta diplomové práce

Student: Kabáč Michal, Bc.
Téma: Multi-modální přepis textu (id 24870)
Oponent: Herout Adam, prof. Ing., Ph.D., UPGM FIT VUT

- 1. Náročnost zadání** **průměrně obtížné zadání**
Úkolem řešitele bylo provést fúzi výsledků existujících rozpoznávačů: rozpoznávače ručně psaného textu a rozpoznávače souvislé řeči. Při řešení bylo třeba nastudovat množství netriviálních znalostí z literatury. Samotné provedení fúze a interpretace výsledků jsou spíše slabší.
- 2. Splnění požadavků zadání** **zadání splněno**
- 3. Rozsah technické zprávy** **splňuje pouze minimální požadavky**
Rozsah textu je podprůměrný, obrázky někdy nejsou příliš přínosné. Algoritmy v kapitole 4 by zasloužily rigoróznější popis (schéma, pseudokód, matematický zápis), který by jistě stál úsilí navíc, ale umožnil by jejich srozumitelnost a výrazně by přidal na rozsahu textu.
- 4. Prezentací úroveň předložené práce** **65 b. (D)**
Struktura a obsah technické zprávy je spíše slabší. Teoretické poznatky v kapitole 2 neukazují na hluboké porozumění existujícím algoritmům a jejich podstatným rysům - řešitel obšírně a formálně definuje konvoluci a vysvětluje pooling... Podrobně vypisuje i principy fungování rozpoznávačů řeči a psaného textu, ale tyto části práce použil více méně jako černé skříňky. Skutečně se věnoval fúzi informací ze dvou rozpoznávačů a v této oblasti je teoretický popis slabý.
Kapitola 4 o implementaci je dosti slabá a působí dojmem hodně rychle sepsaného textu. Hmatatelné objekty v textu (ukázky formátu JSON formou obrázku a screenshot) jsou málo zajímavé. Naopak, algoritmy fúze, které by bylo vhodné zpracovat formou matematických rovnic a algoritmu v pseudokódu jsou popsány souvislou řečí; při jejich složitosti a množství různých parametrů a výpočtů je tento zápis naprosto kryptický a nevhodný (ale řádné a promyšlené sdělení algoritmů by jistě bylo významně pracnější). Kromě toho je nešikovné, že popis algoritmů je v kapitole s názvem "Implementace" - návrh algoritmů je typickou součástí návrhu a ne implementace.
Dost slabé je i experimentální vyhodnocení. Z tab. 5.1 je chybovost rozpoznávače písma 3.02/2.36/2.78, algoritmus 4 vylepšil tyto výsledky pomocí rozpoznávače řeči na 3.03/2.32/2.77, ostatní algoritmy základní výsledek vždy jen zhoršily, někdy i hodně. Navržená a naučená neuronová síť výsledky také ve všech variantách výrazně zhoršila. K nejlepšímu zlepšení tedy došlo v komplexní sadě o 1,7% (0,04 procentního bodu CER), což může být na úrovni šumu měření. Řešitel toto vůbec nereflektuje a nepřemýšlí o tom, v interpretaci výsledků prostě říká, že algoritmus 4 dopadl nejlíp, další algoritmy byly za ním. Nezamýšlí se o dopadech a skutečné interpretaci výsledků, nezamýšlí se o příčinách. Práce nepojednává o tom, že na základě prvních výsledků a jejich interpretace došlo k návrhu jiných algoritmů nebo parametrizace stávajících.
- 5. Formální úprava technické zprávy** **65 b. (D)**
Formální a jazyková úprava technické zprávy je podprůměrná. Text obsahuje jazykové chyby, často přeložené termíny, které by bylo srozumitelnější ponechat anglicky, naopak mnoho obrátů založených na anglickém slovosledu místo slovenského. Formátování je dosti slabé a slabá je kvalita obrázků.
- 6. Práce s literaturou** **75 b. (C)**
Práce cituje velice solidní počet kvalitních publikací. Přesto řešitel v teoretické části práce popisuje spíše banality a text nevypovídá o skutečně hlubokém pochopení problematiky. Zklamáním je provedení a interpretace výsledků - z něho je patrné, že řešitel nemá dobré pochopení vlastností metod a způsobů jejich hodnocení.
- 7. Realizační výstup** **75 b. (C)**
Řešitel vytvořil a adaptoval skripty v jazyce Python.
Zdrojový kód je málo komentovaný a nikdy s použitím docstring, čímž se dost snižuje jeho znovupoužitelnost a komplikuje se možnost navázat na řešitelovu práci.
Vytvořeného kódu a dalších řešení není velice moc - jen co bylo potřeba na prezentované experimenty.
Experimenty technická zpráva prezentuje jako jednorázové: řešitel navrhl čtyři algoritmy a architekturu neuronové sítě a ty vyzkoušel a změřil výsledky. Není patrné, že by probíhal iterativní vývoj, který by z jedněch experimentů učinil závěry a návrhy a postupně návrh řešení vylepšoval a směřoval k dokonalosti.
- 8. Využitelnost výsledků**

Vytvořené řešení nepřineslo zlepšení rozpoznávače textu pomocí rozpoznávače řeči. Nejde ale rychle udělat závěr, že takové vylepšení / fúze není možné - experimenty a jejich interpretace nepůsobí úplně přesvědčivě.

9. Otázky k obhajobě

- Všechny algoritmy vedly ke zhoršení výsledků oproti základnímu rozpoznávači textu. Výjimkou byl algoritmus 4. Jak si to vysvětlujete? Jak byste na základě výsledků navrhl nějaký "algoritmus 5", který by mohl být úspěšnější?
- Vaše neuronová síť ve srovnání s ručně navrženým algoritmem 4 výrazně propadla. V oboru, v němž se pohybujete, je zvykem, že neuronové sítě dosahují lepší úspěšnosti než ručně navržené algoritmy, protože se mohou učit z dat a jejich učení je hluboké. Čím si vysvětlujete tento rozpor?

10. Souhrnné hodnocení

70 b. dobře (C)

Řešitel pracoval na úkolu výzkumného charakteru, musel nastudovat poměrně obtížnou literaturu a problematiku, seznámit se s použitelnými nástroji a hotovými řešeními.

V jednom případě došlo k opravdu drobnému vylepšení (na úrovni chyby měření) základního řešení, v ostatních případech se výsledky výrazně zhoršily. Není takový problém, že výzkumná práce tak asi skončila neúspěchem (to se stává, byť zde bylo lze očekávat nějaký přínos); problém je, že řešitel toto nedokáže interpretovat a analyzovat a přetavit do nějakého závěru či doporučení pro další iterace algoritmů. Mrzuté také je, že samotná technická zpráva nehovoří o žádném iterování řešení - narativ je takový, že řešitel prostě navrhl čtyři algoritmy a neuronovou síť a ty proměřil a čísla vynesl do tabulek.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 31. května 2022

Herout Adam, prof. Ing., Ph.D.
oponent