

Posudek oponenta bakalářské práce

Student: Červíček Petr
Téma: Chatbot postavený na umělých neuronových sítích (id 21988)
Oponent: Novotný Ondřej, Ing., UPGM FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Zadání práce je zcela nepochybně nad rámec bakalářského studia a rozhodně svou náročností převyšuje průměrná zadání. V rámci zadání bylo nutné kromě samotné problematiky chatbotů nastudovat i nemalé množství znalostí z oblasti strojového učení, především pak neuronových sítí. I při využití volně dostupných toolkitů, či modelů, vyžaduje splnění daného úkolu nemalé množství úsilí.
- 2. Splnění požadavků zadání** **zadání splněno s vážnými výhradami**
Vypracovaná práce zadání odpovídá. Veškeré výhrady a jejich závažnost budou diskutovány v závěrečném shrnutí.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
Splňuje v obvyklém rozmezí.
- 4. Prezentací úroveň předložené práce** **59 b. (E)**
Ze začátku je práce logicky strukturovaná. Zásadní problém vidím v popisu teorie neuronových sítí, ta přesahuje z kapitoly 4 do kapitoly 3 (Seq2Seq model) a kapitoly 5 (LSTM). Kapitola 5 by následně mohla být rozdělena mezi kapitolu 4 (teorie) a kapitolu 6 (implementaci-výčet použitých python modulů a toolkitů). Práce se v úvodních částech o historii a dělení chatbotů čte velmi pěkně. U odborných částí textu je to již podstatně horší. Text je občas zamotaný a plný hrubých faktických chyb z oblasti neuronových sítí.
- 5. Formální úprava technické zprávy** **70 b. (C)**
Typograficky a jazykově působí práce průměrně.
- 6. Práce s literaturou** **40 b. (F)**
Autor práce cituje relevantní a poměrně aktuální zdroje. V oblasti oddělení převzatých prvků od vlastní práce byla zjištěna velmi vážná a hrubá pochybení, jenž budou diskutována níže.
- 7. Realizační výstup** **40 b. (F)**
Realizačním výstupem je python implementace chatbota. Kód je v některých ohledech trochu chaoticky členěný. Celková míra úpravy kódu je nevalná. Především absence jakýchkoliv komentářů, dělá implementaci těžce udržitelnou. Zde bylo taktéž zjištěno velmi závažné pochybení s ohledem na autorské právo. Taktéž bude diskutováno níže.
- 8. Využitelnost výsledků**
Práce nepřináší žádné nové výsledky, či výstupy.
- 9. Otázky k obhajobě**
Nejsou.
- 10. Souhrnné hodnocení** **40 b. nevyhovující (F)**
Student v práci cituje bakalářskou práci jenž má stejné zadání a vznikla v loňském roce (Chatbot postavený na umělých neuronových sítích jejímž autorem je Jiří Čechák). Z dané práce přiznává použití trénovacích textových dat. V rámci hodnocení byla mnou prostudována i tato práce. Na základě tohoto, bylo zjištěno, že student využil zdrojových kódů studenta Čecháka. Tato skutečnost nebyla v textu přiznána. Část kódu je místy upravena (odstraněny komentáře, Qt GUI přepsáno na Tk, vlastní definice modelu v kerasu) a drobně reorganizována (možná až nelogicky). Lze očekávat jistou podobnost vzhledem ke stejnému zadání práce, přesto se velmi silně domnívám, že student převzal velkou část kódu. Následně se to reorganizací a drobnými úpravami pokusil skrýt. K domněnce také přispívá zapomenutá hlavička s původním autorem v jednom ze skriptů.

Taktéž bylo zjištěno, že v této práci student z původní práce pana Čecháka převzal modely neuronových sítí, které v jeho práci nemají uvedený zdroj (v práci pana Čecháka je zdroj uveden).

Student byl pozván na osobní konzultaci jeho řešení. Během konzultace byl tázán na teoretické znalosti v oblastech ve kterých se v dokumentaci dopustil hrubých chyb.

- Rozdíl mezi optimalizátorem a objektivní funkcí, či jinou metrikou (V textu se střídavě Adam objevuje jako

- optimalizátor i jako metrika - zaměřuje s klasifikační úspěšností)
- Rozdíl mezi category-cross entropy a accuracy (student popisuje jako velmi podobné jen s rozdílným rozsahem)
- Základní princip a fungování jím použitých modelů (kde je enkodér, kde je dekodér jak probíhá propagace dotazu po jeho transformaci na vektorovou reprezentaci)
- Vysvětlení obecného fungování umělého neuronu (v práci také chybně popsán)

Student nedokázal odpovědět na jedinou otázku, včetně vysvětlení fungování umělého neuronu. Taktéž si ani nebyl vědom ponechané hlavičky s jménem pana Čecháka ve vlastním zdrojovém kódu. Lze se tedy s jistotou domnívat, že nerozumí sám vlastnímu řešení.

Při hodnocení práce bylo bráno v potaz, že se jedná o bakalářskou práci, její účel a předpokládaný výsledek. Bylo přihlédnuto taktéž k tomu, že by student některé zmíněné nedostatky mohl při správném citování legálně použít.

S ohledem na následující:

- Použití dat pana Čecháka (byť řádně citované)
- Použití podstatných částí jeho kódu
- Kopie jeho modelů
- Přítomnost hrubých faktických chyb o NN v dokumentaci
- Absence znalostí, jak o vlastní práci, tak elementárních principů NN

Nevidím jinou možnost než zvolené hodnocení F.

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

V Brně dne: 29. května 2019

.....
podpis