

Posudek oponenta bakalářské práce

Student: Richtarik Lukáš
Téma: Chatbot postavený na umělých neuronových sítích (id 21025)
Oponent: Beneš Karel, Ing., UPGM FIT VUT

1. Náročnost zadání **obtížnější zadání**

Současné state-of-the-art techniky tvorby generativních chatbotů se opírají o modely a postupy strojového učení, které značně přesahují rámec bakalářského studia na FIT. I při využití dostupné implementace modelů je pro dosažení rozumných výsledků třeba nemalého porozumění a rozsáhlé experimentální práce.

2. Splnění požadavků zadání **zadání splněno s drobnými výhradami**

Zadání žádá, aby byla sledována úspěšnost chatbota při jeho postupném zdokonalování. Student však sám věnoval značnou pozornost popisu toho, že metriky, které používá pro sledování úspěšnosti chatbota, jsou do značné míry nevhodné. Z nich jedinou, která popisovanými neduhy netrpí -- pravděpodobnostní objektivní funkci -- nepoužívá. Testování s uživateli proběhlo pouze s jednou verzí chatbota, získané hodnocení není žádným způsobem kalibrováno a tak nebyl do hodnocení úspěšnosti vnesen pořádek.

3. Rozsah technické zprávy **je v obvyklém rozmezí**

4. Prezentací úroveň předložené práce **62 b. (D)**

Na nejvyšší úrovni je práce vhodně strukturovaná, členění do kapitol je přehledné a vhodné. Rovněž rozsah jednotlivých kapitol je vyvážený. V konkrétních aspektech je však zpráva nedostačující k pochopení vykonaných prací:

Použitý dataset, sestavený ze dvou odlišných datasetů, obsahuje více než 200 tisíc trénovacích párů výzva--odpověď, ale při trénování bylo použito nejvýše 80 tisíc párů, většinou dokonce pouze 10 tisíc párů, nadto filtrovaných podle délky. Není proto zřejmé, jaká data byla reálně zahrnuta v experimentech. Navíc není zřejmé, zda byla pro všechny experimenty používána stejná testovací sada a výsledky jsou tedy srovnatelné, zejména při experimentování s BPE a s velikostí trénovacích dat.

Přístup k metrikám je do velké míry řízen implementací, což je na škodu prezentaci. Srovnání různých chatbotů je prezentováno velmi nevhodnou formou, kdy jsou v jednom grafu průběhu výkonu během trénování, na jedné ose, srovnávány různé metriky (slovní přesnost a BLEU). Autor sám dále uvádí, že do výpočtu slovní přesnosti byly zahrnovány i režijní symboly, které de facto nejsou součástí dat.

5. Formální úprava technické zprávy **69 b. (D)**

Celková sazba je v pořádku. Naprostá většina obrázků v práci je bitmapová, a to včetně "obrázků", které sestávají pouze z textu (obr 2.1, 4.1). Vinou přejímání obrázků z různých zdrojů působí kapitola věnovaná neuronovým sítím poněkud neuceleně.

Jazykovou čistotu nemohu podrobně posoudit. V práci však několikrát chybí čárky okolo vět vedlejších. Hláskování anglických výrazů často obsahuje chyby, zvláště výraz myšlenkový vektor (thought vector) se v práci vyskytuje nejméně ve třech formách.

6. Práce s literaturou **62 b. (D)**

Student cituje základní prameny relevantní pro práci. Vlastní odkazy do literatury jsou však často na nepochopitelných místech (před uvedením přejatého poznatku nebo naopak o několik vět či odstavců (!) později).

Odkazy přiznávající původ převzatých obrázků jsou uváděny v soupisu použité literatury, namísto běžnějšího a vhodnějšího uvedení v poznámkách pod čarou.

Obrázek 3.1 je pravděpodobně přejatý (anglické popisy), avšak bez uvedení zdroje.

7. Realizační výstup **65 b. (D)**

Vzniklá sada skriptů má charakter "laboratorního nepořádku". Vzhledem k povaze práce by to samo o sobě nebylo na závadu. Podstatnější je však duplikace dat v kódu (např. překladové tabulky "won't" -> "will not") a používání suboptimálních řešení (lineární vyhledávání ve slovníku, a to i podle číselného indexu; omezené využívání numpy).

8. Využitelnost výsledků

Další využití není zřejmé, po softwarově-inženýrských úpravách by vzniklý balíček skriptů mohl sloužit pro další vývoj generativních chatbotů.

9. Otázky k obhajobě

-

10. Souhrnné hodnocení

62 b. uspokojivě (D)

Nedostatkem této práce je zejména chabé vyhodnocení výsledků. Pro automatické vyhodnocování jsou použity nevhodné metriky a vyhodnocení s lidmi proběhlo pouze s jednou verzí chatbota a je tak neprůkazné. Text práce ukazuje, že si autor osvojil přiměřeně rozsáhlý, avšak poměrně povrchní, soubor znalostí o použitých technikách.

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

V Brně dne: 31. května 2018

.....
podpis