

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Jonáš Kulhánek
Název práce End-to-end dialogue systems with pretrained language models
Rok odevzdání 2021
Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Rudolf Rosa **Role** oponent
Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Téma práce

Student se v práci zabývá velmi aktuálním problémem v oblasti dialogových systémů. V této oblasti můžeme v posledních letech sledovat probíhající a dosud nedokončenou změnu paradigmatu.

Klasické dialogové systémy jsou typicky sestavené z řady samostatných komponent, které jsou do značné míry sestavené ručně, pravidlově, a typicky vyžadují jisté množství dat vytvořených a anotovaných specificky pro tento účel. To sice vede k jejich vysoké přesnosti, ale je velmi náročné je vytvářet či adaptovat na jiné domény či jazyky.

Moderním přístupem pak jsou end-to-end dialogové systémy založené na hlubokých umělých neuronových sítích, které spojují vícero původně samostatných komponent do jedné a obejdou se bez některých explicitních reprezentací odpovídajících jednotlivým mezikrokům původních systémů. V praxi ale moderní systémy zatím obvykle nedosahují úspěšnosti klasických dialogových systémů, a to z mnoha důvodů, z velké části se týkajících nedostatečného množství potřebných trénovacích dat.

Student se v práci věnuje právě moderním end-to-end systémům, přičemž se zaměřuje na několik nedostatků současných systémů a navrhuje řadu vylepšení, která tyto nedostatky částečně řeší a posouvají tak tento obor o něco kupředu.

Obsah práce

Student v práci vychází z existujícího dialogového systému Soloist, který reimplementuje pod názvem AuGPT a obohacuje o některé nové prvky, vesměs vycházející z existující literatury, ale často je v kontextu dialogových systémů užívá zcela nově, či inovativním způsobem. Zásadních zlepšení dosahuje zejména zaměřením se na největší problém moderních dialogových systémů, jímž je nedostatek dat.

Autor jednotlivá navržená vylepšení podrobně vyhodnocuje automaticky i manuálně a svá zjištění dále podrobně diskutuje a interpretuje, čímž prokazuje jejich užitečnost. Interpretace takových zjištění je obvykle poměrně tenkým ledem, neboť lze velmi snadno sklouznout k nedostatečně podloženým spekulacím; s potěšením proto konstatuji, že interpretace, které autor uvádí v textu, jsou obecně dobře podložené a veskrze pravděpodobné. Oceňuji také zahrnutí ukázek výstupů systému. V rámci mezinárodní soutěže DSTC 9 se pak systém AuGPT umístil na třetím místě z 10, což je velmi dobrý výsledek.

Celkově lze říci, že autor se zvládl dobře zorientovat v problematice, identifikoval několik slabých míst současných dialogových systémů, a navrhl řadu vhodných metod zaměřených na zmírnění či odstranění daných slabých míst. Důsledným vyhodnocením navržených metod prokázal úspěšnost většiny z nich, na základě čehož sestavil finální konfiguraci systému, který dosahuje úspěšnosti srovnatelné s nejlepšími světovými dialogovými systémy. Věřím, že předkládaná práce má potenciál ovlivnit oblast výzkumu end-to-end dialogových systémů.

Rád bych také vyzdvihl, že autor výsledky své práce neprezentuje pouze odevzdanou diplomovou prací, ale zveřejnil je zároveň pod svobodnou licencí formou repozitáře na *GitHubu* opatřeného rozumnou dokumentací (zdrojové kódy), položky na *Huggingface* (natrénované modely), a odborného článku na *arXiv.org*, který je zároveň aktuálně v recenzním řízení.

Navrhované metody a srovnání s předchozími pracemi

Za nedostatek práce považuji srovnání s předchozími pracemi, kdy obvykle není z vlastního textu diplomové práce zcela zřejmé, čím se předkládané řešení liší od řešení dříve představených jinými autory. Vlastní metoda je popsána dostatečně, takže studiem a srovnáním citovaných prací lze zjistit, v čem je představovaná metoda nová či inovativní; toto srovnání by ale mělo být již součástí textu předkládané práce. Bez dalšího studia odkazovaných prací je tedy velmi obtížné posoudit skutečný přínos předkládané práce.

Kapitola 3 Related work čítá bez obrázků a jejich popisků jen asi 3 strany textu, přičemž část kapitoly je věnována popisu systémů bez zřejmé vazby na systém navrhovaný autorem (Sequicity, LABES-S2S); zde bych tedy uvítal jasnější zasazení do kontextu. Srovnání s existujícími pracemi pak většinou není přímo součástí této kapitoly, ale je skryté na různých místech vlastní práce. Systém Soloist, ze kterého autor ve své práci vychází, je popsán dobře; tuto část oceňuji.

Autor popisuje pomocnou úlohu (auxiliary task) používanou v systému Soloist, a uvádí, že ve svém systému AuGPT používá jinou pomocnou úlohu, ze stručného popisu v úvodních kapitolách ale není vůbec zřejmé, čím se tyto úlohy liší, naopak se zdá, že jde prakticky o tutéž úlohu; což je také pravda, autorem navržená úloha je pouze rozšířením úlohy užívané v systému Soloist (nicméně velmi zdařilým). Uznávám, že srozumitelně bez dalšího kontextu vysvětlit daný rozdíl je obtížné, přesto by se měl autor pokusit nejpodstatnější rozdíl alespoň naznačit a odkázat na podrobné vysvětlení tohoto rozdílu v sekci 4.7.1.

K dalším navrhovaným pomocným úlohám (4.7.2) pak autor neuvádí žádný odkaz na předchozí práce, lze tedy dovodit, že tyto úlohy jsou vlastním návrhem autora. Chybí mi zde ale jakákoliv rešerše dalších prací, které pracují s pomocnými úlohami v kontextu dialogových systémů, např. ZHAO, Yufan; XU, Can; WU, Wei. *Learning a Simple and Effective Model for Multi-turn Response Generation with Auxiliary Tasks*. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. p. 3472-3483. **Prosím studenta, aby se k tomuto vyjádřil během obhajoby.**

Jen velmi málo jsou popsány předchozí práce zaměřené na augmentaci trénovacích dat pro dialogové systémy pomocí automatického parafrázování (Jin+ 2018, Einolghozati+ 2019), takže není zřejmé, do jaké míry se autorem navržený postup podobá těmto pracem a do jaké míry je nový; přitom právě autorem navržená augmentace dat je asi nejjednodušším přínosem práce! **Prosím studenta, aby se k tomuto vyjádřil během obhajoby.**

Autorem navržená metoda opuštění delexikalizace ve prospěch přímého generování lexikalizovaných výstupů je slibným krokem směrem ke skutečnému end-to-end systému bez nutnosti dalších komponent. Zde ale zcela chybí jakákoliv reference na předchozí práce zabývající se touto myšlenkou a tedy diskuze novosti této myšlenky. Podobnou snahu lze přitom nalézt například již v práci SHARMA, Shikhar, et al. *Natural Language Generation in Dialogue using Lexicalized and Delexicalized Data*. In: *International Conference on Learning Representations: Workshop*. 2017. **Prosím studenta, aby se k tomuto vyjádřil během obhajoby.**

Vůbec v práci není zmíněno, zda a jak již v kontextu dialogových systémů byl využit unlikelihood loss a nucleus sampling; autor mezi přínosy své práce uvádí i zapojení těchto postupů, měl by se tedy výslovně vyjádřit k novosti této myšlenky v daném kontextu. **Prosím studenta, aby se k tomuto vyjádřil během obhajoby.**

Na druhou stranu bych rád ocenil, že autor cituje velké množství odborných prací; bibliografie čítá úctyhodných 10 stran.

Ablační analýza

Ablační analýza patří mezi silné stránky práce. Autor zde vyhodnocuje přínos každé z navržených komponent, přičemž prokazuje, že každá z komponent finální konfigurace systému pozitivně přispívá k celkovému výsledku. Částečně nejednoznačné výsledky pak uspokojivě vysvětluje nedostatky standardních evaluačních metod.

Mám zde ovšem problém s jakousi argumentací kruhem. Autor v textu uvádí, že finální konfigurace AuGPT (případně AuGPT-b) byla zvolena na základě ablační analýzy, která určila, které

metody mají pozitivní efekt a které nikoliv. Samotná ablační analýza ale vychází až z dané finální konfigurace, oproti níž ukazuje pozitivní a negativní přínosy jednotlivých metod. Zdá se tedy, že finální konfigurace byla zvolena na základě jiné, předběžné ablační analýzy, která je poněkud odlišná od té, která je prezentována v práci. Prezentovaná ablační analýza dobře slouží pro ověření, že byla pravděpodobně zvolena optimální konfigurace, ale není pravděpodobné, že tato analýza teprve vedla k výběru finální konfigurace.

Ještě tajemnější je pak volba kontrastivní konfigurace *AuGPT-b*, která je v textu motivována jejími dobrými výsledky, avšak nikde není vysvětleno, co autora motivovalo tuto konfiguraci vůbec evaluovat. Jedná se totiž o konfiguraci, která vynechává dvě z navržených metod: unlikelihood loss, který skutečně dle některých evaluací má negativní přínos (a jeho vynechání je tedy dobře motivované), a filtrování korpusu, které má obecně pozitivní přínos (a tedy motivace jeho vynechání není zřejmá). Nabízí se tedy nutně otázka, zda byly evaluovány i další konfigurace, aniž by toto bylo zahrnuto do textu práce, a co z těchto evaluací vyplynulo.

Prosím studenta, aby se k tomuto vyjádřil během obhajoby.

Text práce

Práce je psaná srozumitelnou angličtinou, avšak obsahuje větší množství drobnějších jazykových chyb, překlepů a neobratných formulací. Sdělení je místy prezentované a strukturované poněkud nelogicky, s užitím nevhodných spojovacích výrazů. Obecně lze říci, že by práci prospěla jazyková korektura. Práce je nicméně i přes tyto mé výhrady bez problémů srozumitelná.

Naopak oceňuji dobrou typografickou úpravu práce.

Závěr

Student ve své diplomové práci úspěšně navrhl, implementoval a vyhodnotil dialogový systém, který je svou úspěšností srovnatelný s nejlepšími světovými systémy. Autor vhodně kombinuje moderní postupy, kterými částečně řeší některé problémy současných end-to-end dialogových systémů. Autor vychází z metod navržených jinými autory, tyto metody dle potřeby vhodně adaptuje, nově používá v kontextu dialogových systémů, či přímo navrhuje jejich alternativy.

Mezi nejsilnější stránky práce patří velmi důsledné vyhodnocení navrhovaného řešení, zahrnující rozmanité automatické i manuální evaluace, přičemž výsledky evaluací jsou dále podrobně kvalitně rozebírány a interpretovány.

Mezi nejslabší stránky práce pak patří zejména nedostatečné srovnání s předchozími pracemi. Tímto nerozporuji vlastní přínos autora, který je nemalý, avšak v textu práce autor jen zřídka důsledně diskutuje, do jaké míry jím navržený postup vychází z existujících prací a čím se od nich liší, tedy co autor navrhuje zcela nově.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 10. 6. 2021

Podpis: