

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Jonáš Kulhánek		
Název práce	End-to-end dialogové systémy s předtrénovanými jazykovými modely		
Rok odevzdání	2021		
Studijní program	Informatika	Studijní obor	Umělá inteligence
Autor posudku	Ondřej Dušek	Role	vedoucí
Pracoviště	Ústav formální a aplikované lingvistiky		

Text posudku:

Shrnutí obsahu Diplomová práce Jonáše Kulhánka se zabývá tzv. end-to-end modely v dialogových systémech – dialogovými systémy reprezentovaným jedinou neuronovou sítí a trénovatelnými vcelku „od začátku do konce“, tj. od uživatelského vstupu až k odpovědi systému. Tento typ architektury dialogových systémů je momentálně v popředí vědeckého výzkumu v oboru a rychle se rozvíjí. End-to-end systémy se rapidně zlepšují a pomalu se začínají vyrovnávat starším modulárním systémům, které používají separátní modely pro porozumění jazyku (zpracování uživatelského vstupu), dialogový management (sledování stavu dialogu a rozhodnutí o další akci systému) a generování odpovědi (vyjádření akce pomocí přirozeného jazyka). Zhruba v posledním roce až dvou se pro end-to-end systémy začaly používat zejména předtrénované jazykové modely založené na neuronové architektuře typu Transformer (BERT, GPT-2 a další). Tyto modely už dříve dosáhly vynikajících výsledků v jiných oblastech zpracování přirozeného jazyka. Pro dialogové systémy taktéž přinesly výrazné zlepšení výsledků a dovolují konstruovat end-to-end systémy schopné obsloužit více domén (témat konverzace).

Datová sada MultiWOZ, kterou práce používá, je dnes nejoblíbenější benchmark pro end-to-end systémy. Jedná se o 10 000 anglických dialogů zahrnujících 7 domén z oblasti turistiky (rezervace a informace o restauracích a hotelích, vlacích, turistických atrakcích apod.). Jednotlivé domény jsou v dialozích často kombinovány a systém tak musí udržovat přehled o preferencích uživatele v několika doménách najednou, podávat požadované informace a ve vhodné chvíli provést rezervaci s požadovanými parametry.

Hlavní cíl diplomové práce – postavit end-to-end systém schopný práce v doménách MultiWOZu – je tedy poměrně ambiciózní. Motivací práce byla také mezinárodní soutěž end-to-end systémů Dialogue Systems Technology Challenge 9 (DSTC9), která se konala na konci loňského roku. Díky ní se systém mohl přímo porovnávat se špičkou v oboru.

Práce měla reimplementovat a pokud možno vylepšit jeden z dosavadních nejlepších end-to-end dialogových modelů, SOLOIST (Peng et al., 2020). Systém AuGPT, který autor implementoval, tedy vychází z modelu SOLOIST a je stejně jako on postaven na předtrénovaném jazykovém modelu GPT-2. Jako většina nejlepších současných end-to-end modelů používá SOLOIST i AuGPT tzv. dvoufázové dekodování: do jazykového modelu se nejprve vloží jako prompt (vstup) předchozí kontext dialogu a jazykový model dekoduje textovou reprezentaci *dialogového stavu* (aktuálních preferencí uživatele). Ty jsou poté použity jako dotaz do databáze, počet výsledků z databáze se vloží jako další prompt jazykovému modelu a ten pak generuje odpověď. V SOLOISTu a dalších modelech

se používá tzv. delexikalizovaná odpověď, kde jsou názvy entit a další údaje z databáze nahrazeny zástupnými znaky, což zjednodušuje generování odpovědi (nahrazení skutečnými údaji se provádí v postprocessingu). Model SOLOIST je trénován jednak na generování výstupu (klasifikace následujícího slova/tokenu) a jednak na klasifikaci konzistence dialogového stavu: dialogový stav je při trénování náhodně nahrazován a systém musí poznat, zda je aktuálně udávaný stav správný.

Autor v práci kompletně reimplementoval základní model SOLOIST, jehož zdrojový kód není veřejný, a dále přidal několik podstatných vylepšení:

- vylepšenou klasifikaci konzistence dialogového stavu, která je pro model výrazně náročnější (model se tak lépe naučí udržovat dialogový stav),
- trénování na datech rozšířených pomocí tzv. zpětného překladač (trénovací data jsou automaticky přeložena do několika jazyků a zpět do angličtiny, takže ve výsledku jsou mnohem variabilnější),
- předtrénování celého modelu na dvou dodatečných datových sadách v podobných doménách,
- použití tzv. *unlikelihood* při generování textu (penalizace opakování slov, aby se zajistila větší variabilita výstupních textů),
- generování plně lexikalizovaných odpovědí, kdy se systém učí kopírovat konkrétní údaje z výsledků databáze.

Výsledný model AuGPT je součtem všech jmenovaných vylepšení.

Všechna vylepšení AuGPT jsou v práci podrobně porovnávána a několika způsoby vyhodnocena ve srovnání se základním modelem SOLOIST i dalším modelům z aktuální literatury. Jedná se o skórování automatickými metrikami na jednotlivých odpovědích a vyhodnocení se simulovaným uživatelem, ale zejména o dialogy s lidskými testery v rámci již zmiňované soutěže DSTC9. Autor dále společně se mnou a dvěma dalšími kolegy provedl 130 testovacích konverzací, při nichž jsme zaznamenávali různé druhy chyb systému, které jsou v práci detailně analyzovány.

Systém AuGPT je sice mírně horší než konkurenční systémy v automatických korpusových metrikách (zaměřených jen na jednotlivé odpovědi), ale překonává je v evaluaci se simulovaným uživatelem, což ukazuje, že si dokáže zachovat konzistentní odpovědi během celého dialogu. Taktéž v soutěži DSTC9 si systém AuGPT vedl výborně – dosáhl 3. místo z 10 účastníků, přičemž jako jeden z pouhých 5 systémů vůbec dokázal porazit velmi silný baseline. Úspěšnost konverzací s AuGPT (bez přihlédnutí k práci s databází) sami testéři v DSTC9 hodnotili jako vůbec ze všech nejlepší. Chování v soutěži DSTC9 i detailní analýza výsledků odhalila některé konkrétní problémy systému (např. občasné „halucinování“ informací, které nejsou v databázi, nebo nezmiňování názvu entity vybrané z databáze). Zároveň však taky ukázala, že systém se v převážné většině konverzací dokáže chovat konzistentně a práci s doménami MultiWOZ zvládá výborně.

Text práce je rozdělen do 8 kapitol. Po stručném úvodu, který uvádí do problematiky a představuje hlavní přínosy práce, vysvětluje 2. kapitola relevantní teoretické pozadí neuronových architektur i dialogových systémů. 3. kapitola podává krátký přehled nejdůležitějších přímo souvisejících prací. V kap. 4 autor představuje architekturu svého modelu AuGPT a vysvětluje všechna implementovaná vylepšení. Kap. 5 popisuje průběh experimentů, kap. 6 shrnuje hlavní výsledky a kap. 7 dále diskutuje některé problémy a možná budoucí řešení. Práce končí krátkým shrnutím.

Průběh prací Hlavní implementační práce na systému probíhaly vloni v létě a na podzim před soutěží DSTC9. Autor ale na základě výsledků soutěže systém ještě během letošního jara dále upravil a vylepšil – zde se jedná zejména o implementaci varianty, která pracuje s plně lexikalizovanými odpověďmi.

Práce probíhala pod mým vedením ve spolupráci se dvěma doktorandy, kteří pomáhali s přípravou na soutěž DSTC9, ale autor měl hlavní iniciativu a sám implementoval celý základní model SOLOIST i všechna vylepšení AuGPT. Stejně tak provádění experimentů vedl autor. Doktorandi implementovali některé podpůrné části kódu, jako např. předzpracování dat. Veškerou práci jsme průběžně konzultovali jak v celé skupině, tak individuálně.

I při sepisování textu diplomové práce jsem s autorem byl v průběžném úzkém kontaktu. Celý text jsme před odevzdáním podrobně probrali. Se spoluprací s autorem jsem plně spokojen.

Hodnocení Práce jednoznačně splnila zadání a převyšuje nároky kladené na diplomovou práci – jedná se o výzkumnou práci na nejvyšší úrovni, která posouvá stav poznání v oblasti end-to-end dialogových systémů. Autor systém AuGPT prezentoval na workshopu k soutěži DSTC9, který se konal v únoru na mezinárodní konferenci AAAI. Na základě výsledků práce vznikl i článek, který se chystáme poslat k recenznímu řízení na jednu z konferencí v oboru.

Předpokládám, že práce se stane základem dalšího výzkumu v této oblasti, ostatně systém AuGPT momentálně začínají používat k dalším experimentům kolegové z Heinrich-Heine Universität v Düsseldorfu, jednoho z předních pracovišť v oboru. Dalšímu výzkumu pomáhá i přehledný zdrojový kód s jednoduchou instalací, který je volně dostupný na GitHubu.

Text práce je psán velmi dobrou angličtinou. Uspořádání do kapitol je systematické a logické; související části na sebe odkazují, což usnadňuje čtení. Veškeré mé připomínky k textu jsme s autorem vyřídili už před odevzdáním a autor mé komentáře zapracoval, takže k textu nemám žádné výhrady.

Celkově práci jednoznačně doporučuji k obhajobě; nemám žádné dotazy.

Práci doporučuji k obhajobě.

Práci navrhuji na zvláštní ocenění.

Myslím, že práce jednoznačně může aspirovat na Cenu děkana za nejlepší diplomovou práci. Důvody jsem popsal již v hodnocení – jedná se vědeckou práci na nejvyšší úrovni, s precizním vyhodnocením, výborným popisem a kvalitním zdrojovým kódem.

V Praze dne 11. 6. 2021

Podpis:

