

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Ondřej Cířka
Název práce Continuous Sentence Representations in Neural Machine Translation
Rok odevzdání 2018
Studijní program Informatika **Studijní obor** Matematická lingvistika

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

Text posudku:

Autor ve své diplomové práci zkoumá reprezentaci vět v neuronových systémech strojového překladu (NMT). Konkrétně se zajímá o reprezentaci vět pomocí vektoru o fixní dimenzi, tak aby bylo možné tuto reprezentaci snadno využít v jiných úlohách jako větný embedink, podobně jako se používají slovní embedinky. Hlavní výzkumnou otázkou je, do jaké hloubky je zachycen význam věty – zda NMT pracuje skutečně s abstraktním významem zdrojové věty a snaží se následně tento vyjádřit v cílovém jazyce, či zda jde o zachycení spíše povrchové, kdy NMT spíše převádí jednotlivá slova a fráze ze zdrojového jazyka do cílového jazyka bez hlubšího porozumění celé větě. Autor v práci poměrně přesvědčivě ukazuje, že se jedná spíše o druhý případ.

Nejlepší současné systémy NMT „bohužel“ nepracují s explicitní reprezentací věty pomocí vektoru fixní dimenze, a tedy neumožňují zamýšlený výzkum snadno provést – věty jsou obvykle v enkodéru reprezentovány sekvencí vektorů (tzv. skrytých stavů), která délkou odpovídá délce věty, a dekodér přistupuje k jednotlivým vektorům pomocí mechanismu atence (attention). Autorovi se povedlo se s tímto problémem poměrně úspěšně vypořádat zavedením modelu složené atence. Klíčovým prvkem jeho postupu je použití vnitřní atence v enkodéru, která nezávisle na dekodéru zakóduje větu do pevného počtu vektorů; konkatenací těchto vektorů tedy získává reprezentaci věty o fixní dimenzi. Dekodér pak nepřistupuje přímo ke skrytým stavům enkodéru, ale pouze k tomuto větnému embedinku. Přes toto podstatné omezení síly systému NMT pozorujeme pouze mírné zhoršení kvality překladu. Toto je prvním silným bodem diplomové práce – autorovi se podařilo úspěšně navrhnout a implementovat alternativní model pro NMT, který má pro účely jeho výzkumu vhodné vlastnosti, a přitom je kvalitou poměrně kompetitivní vůči nejlepším současným systémům.

Pro implementaci dekodéru autor navrhuje použít atenci nad větným embedinkem, což považuji za poněkud překvapivé – atenci chápu jako mechanismus, jak se vypořádat se vstupem variabilní délky, což zde nyní již není třeba, jde tedy jednoznačně o příležitost k použití jiných mechanismů. V experimentech pak autor pracuje i s druhou variantou, kdy v dekodéru atenci vynechá a dekodovacím buňkám dává na vstup přímo větný embedink. V práci však postrádám motivaci či diskuzi nad volbou těchto postupů a zvážení dalších alternativ, což je dle mého názoru škoda. Chápu to tak, že cílem bylo vypořádat se s problémem variabilní délky reprezentace vstupní věty, zatímco pro dekodér byla snaha nic nového nevymýšlet a držet se co nejvíc zavedených postupů (dle mého názoru až do té míry, kdy to nedává zřejmý smysl). Zároveň toto poněkud ředí či zamlžuje přínos autora v této oblasti – zavádí nový mechanismus složené atence, přičemž ale jeho první část (vnitřní atence enkodéru) je přejatá z prací jiných autorů, a druhá část (atentivní dekodér nad větným embedinkem) mi připadá poněkud podivná a není v práci dostatečně teoreticky rozebrána.

Experimentální část práce je velice bohatá. Autor nechal natrénovat větné embedinky pomocí mnoha různých konfigurací hyperparametrů a za použití dvou různých cílových jazyků překladu, což jednak dává práci dostatečnou šíři a důvěryhodnost, a zároveň umožňuje podrobně zkoumat vliv různých nastavení experimentu. Autor následně získané reprezentace vět vyhodnocuje jak přímo v úloze překladu, tak v mnoha dalších úlohách zaměřených na odhadnutí kvality větné reprezentace ve smyslu zachycení významu věty. Experimenty jsou vhodně nastavené, podrobně

motivované a popsané, a jejich výsledky detailně zachycené. Kvalitní experimentální část je další velice silnou stránkou práce.

Slovní vyhodnocení experimentů je často stručnější než bych považoval za ideální. Mnohá pozorování, která lze z výsledků učinit, učiněna nejsou, a čtenáři tak nezbyvá než přímo zkoumat obrovské a ne zcela přehledné tabulky plné čísel. Chápu to tak, že autor se zaměřil pouze na pozorování, která pokládal za důležitá – ta jsou zachycena detailně a často doplněna dalšími velice kvalitními a přehlednými vizualizacemi. Zcela mi ale chybí například rozbor úspěšností dvou variant dekodéru – z výsledků se zdá, že druhá varianta (ATTN-CTX) je přímo v překladu poměrně kompetitivní, a zároveň se takto získané reprezentace zdají být v dalším vyhodnocení obvykle citelně lepší. Očekával bych, že když autor navrhl tyto dvě varianty dekodéru, tak jej bude zajímat, jak si oproti sobě stojí.

Hlavní výsledky jsou ovšem rozebrány velice podrobně a důkladně. Autor porovnává větné embedinky získané jeho metodou s mnoha dalšími embedinky, včetně velice jednoduchých. Oceňuji, že autor porovnal tak velké množství různých reprezentací, což mu zároveň umožňuje dojít k překvapivému a zřetelně průkaznému centrálnímu pozorování celé práce. Autor totiž zjišťuje, že kvalita větných reprezentací ve smyslu zachycení významu věty *negativně* koreluje s jejich kvalitou pro účely strojového překladu: větné reprezentace, které lépe zachycují význam věty, vedou k nižší kvalitě přeladu, a naopak. Toto pozorování je podle všeho nové, a přitom poněkud překvapivé a rozhodně významné.

Autor navíc nezůstává u tohoto pozorování, ale reprezentace dále hlouběji zkoumá, a činí další dílčí pozorování – byť se zde jedná již spíše o hypotézy než jasná fakta, autorova vysvětlení se zdají být rozumná a uvěřitelná. Autor odhaduje, že reprezentace, které jsou v překladu nejúspěšnější, se v zásadě učí lineárně mapovat vstupní slova na výstupní slova, bez hlubšího porozumění významu věty. Ukazuje, že jím použité vnitřní atence se chová spíše pozičně než sémanticky, což dokládá několika vizualizacemi, a zároveň na příkladech jasně ukazuje, proč je toto překážkou v zachycení skutečného významu vět. Až v případech dostatečně jednoduchých reprezentací, které nedokáží snadno zachytit informace pozičním způsobem, je systém „donucen“ zachytit v embedinku význam věty. Kvalita reprezentace tím stoupá, ale kvalita překladu klesá. Autor tedy uzavírá, že úloha strojového překladu se nezdá být dobrým způsobem pro získávání reprezentací vět zachycujících jejich význam. Přestože jím navržená metoda v určitých konfiguracích zachycuje význam věty relativně dobře, prakticky ve všech experimentech je význam ještě lépe zachycen prostým průměrem standardních slovních embedinků. Přitom průměrovat slovní embedinky je metoda až trapně triviální, se zřejmými problémy (například nemožnost zachycení slovosledu), dobře prozkoumaná a datově i výpočetně nenáročná; přesto autorovy experimenty ukazují, že je velmi obtížné ji překonat pomocí výrazně sofistikovanějších metod NMT, u kterých by bylo možné očekávat, že díky náročnosti úlohy překladu budou inherentně dobře zachycovat význam vět. Zde je mi poněkud líto, že autor detailně zkoumá pouze reprezentace vět v NMT, zatímco průměrování embedinků pouze vyhodnocuje v některých tabulkách a dál se mu příliš nevěnuje. Chápu to ale tak, že toto bere jako relativně prozkoumanou záležitost, která navíc není přímo tématem práce – tím je zkoumání větných reprezentací v NMT.

Práci považuji za velmi kvalitní, zejména díky aktuálnosti tématu i postupu, úspěšnému návrhu alternativního enkodéru pro NMT, a bohaté a důsledné experimentální části. Oceňuji širokou škálu metod pro vyhodnocení kvality větných reprezentací; autor trefně poznamenává, že žádná z nich není ideální, a správně se proto rozhoduje je využít všechny. Pozorování, která autor v práci činí, pak považuji za velmi zajímavá a významná. Práce je vhodně strukturovaná a je psaná dobrou angličtinou, s naprosto ojedinělými jazykovými či typografickými chybami a překlepy, které, až na popis tabulky 4.1, nebrání porozumění. Oceňuji také vizualizace v předposlední kapitole práce, které jsou velmi dobře vyvedené, hezké, přehledné, a hlavně informačně bohaté – nejsou v žádném případě samoúčelné, naopak poskytují vhled do vnitřní struktury modelů, který je zcela zásadní pro učinění některých pozorování a hypotéz. Například vizualizace složené atence jsou pro mě velmi překvapivé – neočekával bych, že se budou chovat tak silně pozičně.

Za nedostatek práce považuji zpracování souvisejících prací. Autor důsledně cituje mnohé relevantní práce, což ukazuje jeho dobrou orientaci v oboru, a poukazuje na některé zajímavé souvislosti. Oceňuji, že detailně rozebírá vše, co přímo využívá či na co navazuje ve své práci, i

několik souvisejících konceptů, na které přímo nenavazuje, jako například alternativní teoretické i praktické způsoby reprezentace významu věty. Mnohé související práce ale zmiňuje tak stručně, že citace nemá větší význam než upozornění čtenáře na existenci dané práce, ale více informace nenese. Autor zmiňuje například několik prací, které také zkoumají vnitřní reprezentace v neuronových sítích, již ale neříká, k jakým zjištěním tito autoři došli, a do jaké míry jsou jejich pozorování v souladu či v opozici s pozorováními v této diplomové práci, ani jak se jejich postupy liší či podobají. Tento příliš stručný způsob zmiňování souvisejících prací se přitom objevuje na několika místech.

Nelze také nezmínit poměrně nekvalitní tisk a vazbu práce. Listy s ilustracemi jsou tištěny kvalitně, tisk listů bez ilustrací je ale bledý a nekvalitní, s nízkým rozlišením. Kvůli oboustrannému tisku je pak práce příliš tenká, takže se ve vazbě vlní (anebo je vazba prostě nekvalitní).

Rád bych, aby se autor vyjádřil k následujícím otázkám či problémům:

1. Motivace pro implementaci dekodéru pomocí atence, diskuze alternativ.
2. Autor zmiňuje větší variabilitu a méně spolehlivé závěry u experimentů s němčinou (jak výsledky jednotlivých experimentů, tak jejich korelace s kvalitou překladu). Zajímalo by mě, do jaké míry je to dáno prostě tím, že experimentů na němčině bylo provedeno víc než na češtině. Změnila by se situace, pokud bychom pro němčinu uvážili pouze ty experimenty, které byly provedeny i na češtině?
3. Autor uzavírá, že reprezentace vět v NMT dostatečně nezachycuje význam věty; význam je lépe zachycen i v obyčejném průměru slovních embedinků. Zajímalo by mě, zda autor vidí nějakou jinou úlohu, kde by očekával, že vnitřní reprezentace vět budou jejich význam zachycovat lépe.
4. Autor v závěru naznačuje, že nedostatečné zachycení významu věty v NMT je dáno způsobem trénování, kdy dle jeho názoru trénování systému k tomu, aby dosáhl co nejvyššího BLEU či co nejnižší křížové entropie za použití pouhé jedné reference, nemotivuje systém k tomu, aby pracoval s významem věty. To je samozřejmě problém, pokud chceme interní reprezentaci věty z NMT využít jinde. Není mi ale jasné, zda to je dle autora problematické i přímo pro NMT. Pokud ano, pak by mě zajímalo, v čem spatřuje hlavní problém, a případně jakým způsobem si myslí, že by mohlo být možné jej řešit.

Práci doporučuji k obhajobě.

Práci navrhuji na zvláštní ocenění.

Diplomovou práci jednoznačně považuji za nadprůměrně kvalitní. Autor se zabývá velice aktuálními tématy, pracuje s nejlepšími současnými systémy, navrhuje vlastní kompetitivní variaci na tyto systémy, kvalitně provádí a vyhodnocuje velké množství relevantních experimentů, a dochází k překvapivým zjištěním. Jeho centrální pozorování je nové a poměrně zásadní, a přitom díky důsledné experimentální části průkazné a důvěryhodné. Kvalitu práce dokládá i publikace přijatá na nejlepší mezinárodní oborovou konferenci (ACL), což je v případě magisterského studenta obrovským úspěchem. Svým významem tato diplomová práce dle mého názoru předčí i některé práce dizertační.

V Berouně dne 9. 6. 2018

Podpis: