

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Iryna Tryhubyshyn
Název práce Mutual Relation of Machine Translation and Quality Estimation
Rok odevzdání 2022
Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Mgr. Aleš Tamchyna, Ph.D. **Role** vedoucí
Pracoviště Memsources, a.s.

Text posudku:

Diplomová práce Iryny Tryhubyshyn se zabývá vztahem mezi systémy pro strojový překlad a systémy pro automatický odhad kvality strojového překladu (quality estimation, QE). Hlavním úkolem práce je empirický průzkum souvislosti mezi silou modelu pro strojový překlad (měřeno velikostí trénovacích dat, množstvím parametrů atp.) a schopností různě silných modelů pro QE tyto překlady úspěšně hodnotit. Kromě těchto základních experimentů má práce prozkoumat také vliv domény textu na úspěšnost QE, a konečně také schopnost natrénovaných modelů pro QE hodnotit existující akademické a komerční překladače.

Práce je poměrně přehledně strukturována, text je psán anglicky. Jazyková úroveň práce je spíše průměrná s relativně častými gramatickými chybami a neobratnostmi. Autorka se vyjadřuje velmi úsporně, někdy až na úkor srozumitelnosti.

V teoretické části se autorka věnuje jak vysvětlení základních pojmů relevantních pro dané téma, tak hlubšímu průzkumu existující literatury. Autorka navrhuje jako hlavní faktory ovlivňující sílu strojových překladačů množství trénovacích dat a velikost modelu. S využitím řady existujících výzkumů podrobně zdůvodňuje vhodnost této volby. Podobný přístup pak navrhuje využít i při srovnávání síly systémů pro odhad kvality.

Teoretická část pokrývá veškerou relevantní literaturu, autorka navíc uvádí řadu podnětných informací a ukazuje zajímavé souvislosti mezi jednotlivými články. Ačkoliv některé pasáže působí z pohledu práce redundantně, např. podrobná definice tzv. attention mechanism v architektuře Transformer, celkově autorka prokazuje dobré porozumění problematice strojového překladu, evaluace a odhadu kvality.

Hlavním příspěvkem práce je experimentální část. Rozsah provedených experimentů je dostatečný a zcela splňuje zadání práce. Ocenil bych ale podrobnější rozbor a snahu o vysvětlení získaných výsledků. Zvolená vizuální forma prezentace výstupů se mi navíc jeví jako poněkud nešťastná, základní zjištění z grafů lze vyčíst jen poměrně obtížně. Celkově tato část práce působí bohužel poněkud nepřehledně a příliš stručně; věřím, že podrobnějším průzkumem některých výsledků by bývalo bylo možné dojít k zajímavým závěrům.

Naopak oceňuji, že kromě architektury BiRNN byly provedeny také experimenty s modernějším přístupem Predictor-Estimator v kombinaci s předtrénovanými modely XLM-R. Tyto výsledky v práci dodávají důležitý kontext ostatním zjištěním.

S ohledem na dílčí výsledky by také bylo zajímavé provést experimenty s modely QE, které

by při trénování využívaly data ze strojových překladačů různé kvality. Podobně zajímavé by z mého pohledu bylo prozkoumat, jak výsledky ovlivňuje volba cílové metriky (v práci se používá výhradně TER). S ohledem na rozsah již provedených experimentů ale nelze tyto návrhy vnímat jako kritiku, spíše jako doporučení pro navazující výzkum.

Součástí práce je příloha se zdrojovými kódy a konfiguračními soubory, pomocí kterých lze experimenty reprodukovat. Kód je dobře srozumitelný a obsah je přehledně strukturován.

Celkově se z mého pohledu jedná o poměrně zdařilou práci, hlavní prostor pro zlepšení vidím v jazykové úrovni a v analýze výsledků. Práci doporučuji k obhajobě.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Datum 31. 8. 2022

Podpis