

Stanovisko vedoucího disertační práce

Jméno a příjmení autora posudku: Jan Hajič

Jméno a příjmení autora práce: Jindřich Helcl

Název práce: Non-Autoregressive Neural Machine Translation

Vlastní text (sem prosím napište text posudku, délka textu posudku není omezena):

Obsah práce

Disertační práce Jindřicha Helcla je na téma strojového překladu, konkrétně se zaměřuje na posouzení metody tzv. neautoregresního strojového překladu (tj. strojového překladu pomocí neautoregresního umělých neuronových sítí).

Po úvodní části autor popisuje celkové současné schéma automatického překladu pomocí umělých neuronových sítí (Neural Machine Translation) a popisuje jeho principy a relevantní součásti. Speciální pozornost věnuje autoregresnímu dekodéru, jako nutnou podmínku porozumění detailnímu popisu neautoregresního modelu, kterému jako hlavnímu tématu disertace věnuje celou třetí kapitolu, v jejímž rámci popisuje i varianty tohoto modelu nebo modely příbuzné (např. semi-regressive MT). Popisuje i motivaci pro vznik neautoregresních modelů, tj. rychlost dekodování. V další kapitole pak popisuje tu část neautoregresního modelu, které se speciálně věnoval, a to „Connectionist Temporal Classification (CTC)“, jako možné metodě zlepšení plynulosti překladu, který jinak trpí nezávislým generováním jednotlivých slov v překladu.

V páté, klíčové kapitole, popisuje experimenty, které prováděl s cílem prozkoumat neautoregresní modely ve srovnání s modely autoregresními a porovnat jejich chování jak v oblasti přesnosti a obecně kvality překladu, tak z hlediska rychlosti, kterou předchozí publikace považovaly za hlavní přínos těchto metod. Výsledky experimentů z hlediska kvality a času dekodování (inference) uvádí v závěru této páté kapitoly.

Poslední, šestá kapitola obsahuje zobecněné závěry práce.

Hodnocení

Provedené experimenty a jejich popis dodržují veškeré standardy obvyklé v experimentálních tématech a úkolech v oblasti zpracování přirozeného jazyka (NLP). Experimenty jsou provedeny rigorózně, na datech, která jsou všeobecně k dispozici pro možné ověření a replikace experimentů, a velmi bohaté srovnání s předchozími experimenty a pracemi zabývajícími se stejným tématem a metodou je provedeno rovněž standardně a objektivně. Experimentů bylo navíc provedeno velké množství, a to jak experimentů replikačních, tak experimentů s autorem modifikovanými modely (knowledge distillation). Všechna porovnání byla provedena rigorózně a podle současných standardů. I když autor nedošel k lepším výsledkům než pro model autoregresní, jsou jeho výsledky velmi cenné, protože šel za rámec experimentů a porovnání obvyklých pro experimenty s neautoregresními modely – tj. používal novější data (než jen WMT 14 data, která jsou obvyklým, ale často jediným standardem), použil silnější

„baseline“ než předchozí práce apod. Je samozřejmě nemožné prohlásit, že tím prokázal, že neautoregresní modely již nelze zlepšit nebo že jsou inherentně horší než modely autoregresní, ale jeho výsledky to silně naznačují. V závěrech autor pak navrhuje, jak postupovat pro „férové“ porovnávání výsledků používajících neautoregresní metody s metodami ostatními, a to jak z hlediska použití testovacích datových sad a z hlediska porovnání efektivity (tj. zejména rychlosti inference/dekódování). Tyto závěry považuji za zcela legitimní.

Doporučení k obhajobě:

Předkládaná disertační práce splňuje minimální nároky kladené na disertační práci a jako takovou ji doporučuji k obhajobě.

V Praze dne: 3. 2. 2022

Podpis: Jan Hajič, v. r.