

Posudek disertační práce

Posudek vedoucího

Řešitel: Ing. Tom Kocmi

Název práce: Exploring Benefits of Transfer Learning in Neural Machine Translation

Vedoucí: doc. RNDr. Ondřej Bojar, Ph.D.

Tom Kocmi se ve své disertační práci zabývá neuronovým strojovým překladem (NMT), konkrétně otázkám spojeným s metodami trénování velkých neuronových modelů a možnostmi využití ne zcela vhodných trénovacích dat nebo modelů získaných dříve. Souhrnně takové metody spadají pod termín „transfer learning“, tj. techniky učení, kde dochází k přenosu znalostí z jedné oblasti do jiné nebo z jednoho natrénovaného modelu do druhého.

Text práce je přehledně organizován do šesti kapitol. Po stručném úvodu se autor podrobně věnuje základním technickým pilířům, na nichž práce stojí (kap. 2 Background), ať už jde o relevantní zdroje jazykových dat, konkrétní použitá trénovací data nebo o metody vyhodnocování kvality strojového překladu. V této kapitole bychom neměli přehlédnout popis autorova systému LanideNN pro identifikaci jazyka textu na úrovni jednotlivých slov nebo teoreticky i menších jednotek. Následuje kap. 3 shrnující současný stav techniky neuronového strojového překladu (slovní embeddingy, podslovní jednotky, architektury neuronových sítí, kritérium konvergence ad.).

První díl těžiště práce nalezneme v kap. 4 Transfer learning. Práce odlišuje metody, které mají přístup k „otcovskému“ modelu, tj. modelu jehož dovednosti je třeba využít v kýženém „potomkovi“, a metody, kde je otcovský model fixní, a k němuž ani nemusíme mít původní trénovací data. V obou případech Tom pečlivě studuje několik možných variant metody a hledá za pomoci automatického měření kvality překladu ty empiricky nejlepší. Souhrnně lze říci, že většina zkoumaných konkrétních technik pomáhá a zlepšení jsou sice nejvýraznější, pokud má potomek málo vlastních trénovacích dat, ale přínos je patrný vždy.

Druhý díl těžiště představuje kap. 5 Analysis, která experimenty z předešlé kapitoly podrobuje důkladnému rozboru a doplňuje různé kontrastivní pokusy. Výsledkem tohoto postupu je úctyhodná řádka explicitně zapsaných pozorování, vždy podložených experimentálními výsledky. Například se dozvídáme, že z otcovského modelu se příliš nepřenáší (ne)schopnost generovat věty správných délek: otcovský model omezený na trénovací věty pouze určitého rozsahu délek není schopen reagovat adekvátně na nové vstupy a generovat věty jiných délek, než jaké viděl při trénování. Toto omezení ale nepoškozuje výrazně schopnost otcovského modelu pomoci v transferu potomkovi. Podobně je tomu v oblasti slovosledných vzorců: trénujeme-li otcovský model na větách se zamíchanými slovy na zdrojové nebo cílové straně, výrazně to jeho výkon poškodí; i tak však takto poškozený model poměrně dobře slouží v transferu.

Závěrečná kapitola kromě souhrnu přináší odhad ekologické stopy výpočtů provedených pro účely disertace. Jako vedoucí jsem toto hledisko dosud vždy přehlížel a zajímal se pouze o horní mez nákladů: pokud bylo něco s naší výpočetní silou dosažitelné, bylo žádoucí pokusy

provést. Ekologická stopa správným způsobem apeluje na efektivitu experimentů. Příznivé je pak Tomovo zjištění, že spotřebovaná energie je v rozumné relaci se spotřebou při cestách na konferenci, kde byly dílčí výsledky prezentovány.

Práce je psána dobrou angličtinou, i když občasné gramatické chyby i přes velkou snahu do finální verze unikly. Z ostatních formálních hledisek je práce dle mého soudu vynikající, ať již se jedná o sazbu nebo práci s literaturou, citování vlastních i cizích prací.

Na práci Toma Kocmiho si cením samostatného tvůrčího přístupu po celou dobu studia i při psaní samotného textu disertace. Nadstandardní byla a je Tomova schopnost práce s termíny a včasným dokončováním úkolů. Vedlejším výsledkem dobře řízené pozornosti je celkem 17 vědeckých článků, do nichž Tom během svého doktorského studia větší či menší měrou přispěl, včetně publikací v hlavních sekcích významných konferencí (EACL, WMT, RANLP, TSD ad.).

Největší přínos v předkládané práci spatřuji v přehledné prezentaci velmi širokého spektra pozorování. Jak autor správně upozorňuje, tato pozorování nejsou plně zobecnitelná, vždy je třeba mít na paměti konkrétní podmínky pokusu (jazyky, použitá data, ad.). Tato pozorování vesměs dokládají, že na různých snadno pojmenovatelných rysech otcovského modelu příliš nezáleží. I z poškozeného otcovského modelu (jen věty učitých délek ap.) nebo otcovského modelu pro nesouvisející jazyky lze úspěšně přenést nějaké „znalosti“. Největší zisk z přenosu nastává při otcovském modelu trénovaném co nejdéle na co největších datech. Rád bych v této souvislosti položil doplňující otázku: Jak by autor vlastními slovy shrnul, co je tedy nejspíš ta přenášená znalost?

Závěrem jednoznačně konstatuji, že předložená práce Ing. Toma Kocmiho splňuje nároky kladené na disertační práce, a proto ji doporučuji k přijetí.

Praha, 28. listopad, 2019.

doc. RNDr. Ondřej Bojar, Ph.D.