

Oponentský posudek diplomové práce

Bc. Petra Vysušilová: Czech NLP with Contextualied Embeddings

Předložená diplomová práce se zabývá aktuálním problémem analýzy přirozeného jazyka za použití nejmodernější technologie kontextových embeddingů, tzv. jazykového modelu BERT (a jeho variant).

Konkrétní úlohy NLP řešené v práci s pomocí BERT jsou tři: POS tagging, lematizace a analýza sentimentu (na několika různých souborech dat). Práce (psaná anglicky) je rozdělena do tří kapitol a dalších obvyklých částí (závěr, reference, seznamy). V první kapitole jsou popsány obecné úlohy zpracování přirozeného jazyka, dále pak základní aspekty hlubokého strojového učení pomocí umělých neuronových sítí, a speciálně pak modely typu BERT. Ve druhé kapitole jsou popsány provedené experimenty; je popsán způsob trénování (nastavení) hyperparametrů, a poté společně úloha lematizace a POS taggingu. V závěru druhé kapitoly je popsána třetí úloha, tj. analýza sentimentu na třech datasetech. Třetí kapitola popisuje implementaci a ukazuje ji na příkladu. Práce je ukončena závěrem, seznamem literatury, seznamem obrázků, seznamem tabulek a obsahuje rovněž tři přílohy: o learning rate, analýzu tagů (značek), na nichž se dosáhlo největšího zlepšení při POS taggingu, a analýzu vůči modelu mBERT. K dispozici jsem měl i errata dodaná po odevzdání práce.

Hodnocení

Práce je psaná srozumitelnou a slušnou angličtinou, i přes drobné neobratnosti. Začátek (první kapitola), kde se popisují v práci řešené úlohy NLP, a základní informace o hlubokém učení a modelech BERT, je velmi pěkně napsaná rozsahem i obsahem, s dostatečným matematickým, grafickým a obrázkovým doprovodem a citacemi relevantní literatury. Popis experimentů ve druhé kapitole je naopak velmi stručný, a jeho část je věnována předchozím pracím jiných autorů, popisu struktury dat a morfologie; to je zcela v pořádku (vše je opatřeno vhodnými a bohatými citacemi), nicméně vlastní experimenty autorky jsou popsány velmi stručně a do jisté míry implicitně (podle definici popisů v tabulkách). Podstatné informace v textu ovšem nechybí a jsou i graficky zdokumentovány (např. obr. 2.2, architektura joint modelu). Analýza výsledků pro tagging a lematizaci je velmi stručná, cenné jsou však tabulky v appendixu 2 a 3 (kde je zřetelně vidět, že došlo ke zlepšení dlouhodobého problému s rozhodováním mezi 1. a 4. pádem v češtině, jak se ostatně píše i textu na str. 50). Výsledné hodnoty evaluace jsou velmi dobré, a překonávají SOTA výsledky. Pro analýzu sentimentu byly použity tři české datasety. V práci jsou podrobně popsány, a jsou popsány i výsledky na datasetech s použitím různých jazykových modelů, a jsou dále porovnány proti baseline (Naive Bayes). Výsledky jsou opět velmi dobré a zlepšují SOTA na dvou ze tří použitých datasetů. Závěrečná kapitola nazvaná „Implementation analysis“ je pomůckou pro reprodukovatelnost výsledků (kód je volně k dispozici, stejně jako použité datasety u

všech úloh), nikoli však skutečnou analýzou. Závěr je stručný a ukazuje některé směry dalšího možného postupu.

Závěr: kladně hodnotím dosažení SOTA výsledků, velmi dobře popsany úvod (kap. 1), množství provedených experimentů a skutečně rozsáhlou a relevantní literaturu, která může sloužit jako základ pro autory podobných DP v této oblasti. **Negativní** je naopak velká stručnost při popisu experimentů a kódu (kap. 3 se o to sice snaží, ale replikovat popsané experimenty bude do značné míry stejně vyžadovat trial and error postup, i díky – nutno říci, že bohužel v DP obvyklé absenci dostatečných komentářů v kódu). Analýza chyb z hlediska úloh jako taktových (tagging, lematizace, analýza sentimentu) je zcela minimální. Přes uvedené výhrady práci **doporučuji** k obhajobě.

Praha, 13. 8. 2021

Jan Hajič, ÚFAL MFF UK