
Hodnocení vedoucího bakalářské práce

Jakub Šmíd

Detekce polarity textu s využitím mezijazyčné transformace

Obsah Práce

Práce se zabývá detekcí polarity českého textu s využitím dat z jiného jazyka (angličtiny) a je spíše výzkumného charakteru. Cílem práce bylo ověřit možnosti použití mezijazyčných transformací, které dovolují vytvořit systém pro detekci polarity českého textu a to pouze při použití dat anotovaných v anglickém jazyce.

V první části práce jsou spolu s vybranými mezijazyčnými transformacemi popsány metody a algoritmy strojového učení použitelné pro detekci polarity textu. Jejich plné pochopení je pro bakalářského studenta poměrně náročné. V druhé části práce student nejdříve vytvořil a otestoval modely pouze s českými daty (tzv. monolingvální modely). Dále student zkoumal dvě mezijazyčné transformace (metodu nejmenších čtverců a ortogonální transformaci), které transformují sémantické prostory dvou jazyků do jednoho sdíleného bilingválního prostoru.

S využitím těchto transformací student natrénoval mezijazyčné modely pouze s anglickými daty, které jsou schopny predikovat polaritu českého textu (i když nebyla použita žádná česká anotovaná data) tzv. *zero-shot* klasifikace. Nad vytvořenými modely jsou provedeny experimenty na českém a dvou anglických datasetech. Výsledky experimentů jsou porovnány s výsledky monolingválních modelů a ukazují, že je možné dosáhnout poměrně dobré úspěšnosti, která je pouze o 5 % až 6 % horší než při použití českých dat. Může se zdát, že 5 % je poměrně velký rozdíl, ale v kontextu této práce se jedná o dobrý výsledek.

Původnost práce a kvalita řešení

Práce je zcela původní, student dokonce jako první¹ aplikoval mezijazyčné transformace pro detekci polarity v českém jazyce. Student používá standardní metody pro klasifikaci textu.

Formální úroveň práce je v pořádku. Dokument bakalářské práce je vysázen v \LaTeX a je použita vektorová grafika. Práce je implementována v jazyce Python a jsou použity standardní knihovny pro neuronové sítě. Zdrojový kód je přehledný, je vhodně rozdělen do tříd a samostatných souborů a bylo možné jej bez problému spustit.

Využitelnost dosažených výsledků

V práci bylo ukázáno, že je možné použít mezijazyčné transformace a pouze anglická data pro detekci polarity v českém textu s poměrně dobrou úspěšností. Získané výsledky mohou posloužit jako základ pro další výzkum těchto metod, u kterého se předpokládá výstup ve formě publikace.

Spolupráce s vedoucím a aktivita studenta

Spolupráce probíhala naprosto bezproblémově. Student se pravidelně účastnil konzultací a vždy včas plnil dohodnuté úkoly. Student se aktivně snažil vnést do práce své vlastní nápady potřebné pro řešení problémů, které se v průběhu práce objevily.

¹Není mi známá žádná publikace, která by aplikovala mezijazyčné transformace na detekci polarity českého textu.

Splnění zadání

Zadání bylo kompletně splněno. S prací jsem byl nad očekávání spokojen. Práce se svým rozsahem, kvalitou a obtížností daného tématu blíží spíše práci diplomové než bakalářské.

Navrhuji hodnocení známkou **v ý b o r n ě** a práci doporučuji k obhajobě.

V Plzni dne 31. května 2021

Ing. Pavel Přibán
(vedoucí BP)