

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Patrícia Březinová
Název práce Computational analysis and synthesis of song lyrics
Rok odevzdání 2021
Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Rudolf Rosa **Role** oponent
Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

K hlavním cílům práce

Tématem práce je automatická analýza a syntéza písňových textů; zadání pak jako podstatnou součást zmiňuje metody automatického vyhodnocování kvality textu.

V diplomové práci se autorka zaměřuje právě na automatickou analýzu či evaluaci. Široké téma automatického vyhodnocení kvality písňového textu se ovšem v práci zúžilo prakticky výhradně na detekci a evaluaci *rýmů*: určení veršů, které se rýmují, a stanovení schématu rýmu. Výsledek této analýzy je pak vizualizován ve webovém rozhraní a vypočítáno skóre, které vyjadřuje, jak dobře se text rýmuje. Detekce rýmů se zdá být vyřešena poměrně dobře, vyvinutý nástroj *Rhyme Detector* má relativně vysokou přesnost, jak dokládá evaluace v práci a jak jsem i měl možnost si ověřit v aplikaci, která je součástí práce. Vizualizace je také provedena dobře a může být užitečným nástrojem například pro literární vědce.

Zajímavé je i použití detekce rýmů a automatického skóre pro analýzu písňových textů v korpusu a následný podrobný rozbor získaných výsledků a porovnání jednotlivých žánrů na základě této analýzy. Ze své pozice nedokážu zodpovědně posoudit novost a přínosnost této analýzy, a ani autorka zde nijak neporovnává svá zjištění s existující literaturou, ale předpokládám, že některé zde uvedené poznatky mohou být nové a zajímavé v rámci kvantitativní literární vědy.

Skóre pro hodnocení kvality písňového textu je ovšem spočítáno poměrně přímočaře a hodnotí v zásadě jen kvalitu rýmování. V práci není podrobně diskutováno, do jaké míry toto skóre postačuje pro obecnou úlohu hodnocení kvality písňového textu, a jaké další aspekty by bylo vhodné v automatické evaluaci vzít v potaz (částečně je toto diskutováno v závěru práce). Skóre není ani v tomto smyslu nijak evaluováno, například formou korelace tohoto skóre s lidským hodnocením kvality textu. Není tedy zřejmé, jak úspěšně by toto skóre mohlo plnit svůj původně zamýšlený účel, tedy porovnání kvality jednotlivých automaticky vygenerovaných písňových textů.

Prakticky neřešeno zůstává druhé z hlavních témat zadání, tedy generování písňových textů. Autorka tomuto tématu věnuje pouze dvě stránky (kapitola 6), přičemž zde pouze triviálním způsobem využívá existující nástroj (GPT-2) a demonstruje několik získaných výstupů. Analýza těchto výstupů je velmi stručná a pouze ruční, autorka tedy nijak nevyužívá nástroje pro analýzu písňových textů vyvinuté v této práci – přičemž jejich využití pro automatické generování písňových textů bylo zjevně původní motivací pro zahrnutí tohoto tématu do zadání práce. Tuto část zadání tedy hodnotím jako nesplněnou (byť práce obsahuje rudimentární náznak řešení).

K jednotlivým částem práce

Za hlavní část práce lze jednoznačně považovat nástroj *Rhyme Detector*, který v písňových textech vyhledává rýmy, určuje jejich typy, a přiřazuje textu schéma rýmů. Autorka zde vhodně vychází z existující literatury (zejména Plecháč 2017 a 2018) a využívá dostupná existující data a nástroje. V praxi se ovšem ukazuje, že některé problémy byly dosud řešené nedostatečně či zcela neřešené, autorka zde tedy navrhuje a implementuje vlastní řešení či vylepšení. Výsledné řešení je poměrně komplexní a kombinuje mnoho nástrojů a postupů, včetně strojového učení. Zároveň z textu práce vyplývá, že během vývoje autorka vyzkoušela i mnohé další postupy, podrobně zde diskutuje

postupy navrhované v literatuře, porovnává jednotlivé postupy, které implementovala, a tedy poměrně důsledně empiricky zdůvodňuje většinu rozhodnutí ohledně finální volby metody i jejích parametrů, a poctivě diskutuje i její nedostatky. Vzniklý nástroj pak důsledně evaluuje a porovnává s existujícím nástrojem RhymeTagger, přičemž ukazuje, že jí vyvinutý nástroj dosahuje srovnatelné úspěšnosti. Tuto část práce tedy hodnotím veskrze kladně, jde o kvalitně provedený výzkum a vývoj; obzvláště pak oceňuji velkou šíři postupů, které autorka pro řešení této úlohy vyzkoušela.

S detektorem rýmů souvisí i vizualizační nástroj ve formě webové aplikace, který příjemným interaktivním způsobem umožňuje uživateli provést analýzu rýmů ve vloženém písňovém textu. Vizualizaci považuji taktéž za velmi dobře provedenou, díky ní je možné vytěžit maximum z vyvinutého Rhyme Detectoru, který tak může být užitečný zejména pro usnadnění analýzy a práce s existujícími písňovými texty, například pro provádění rozborů rýmů na větším množství textů. Vizualizační nástroj z mého pohledu znamená odchýlení se od původního směřování práce, avšak vzhledem k praktické absenci původně zamýšlené druhé části práce věnované generování textů (a tedy nevyužití automatických analýz pro tento účel) dává vizualizace nový smysl detektoru rýmů a tedy i celé práci.

Obecně bych tedy rád ocenil, že v situaci, kdy se ukázalo, že původní cíl práce se nepodaří splnit, se povedlo stanovit si nový cíl, který byl následně splněn. Jestliže tedy hrozilo, že vznikne nedokončená a nesmyslná práce, je nutné ocenit, že posunem ve směřování práce se podařilo vytvořit práci sice částečně jinou a skromnější než by odpovídalo původnímu zadání, avšak jednoznačně smyslupnou a ucelenou.

Text práce má různorodou, ale poměrně dobrou kvalitu. Zvláště oceňuji kapitolu 1, která na 12 stránkách podrobně rešeršuje existující postupy a nástroje; s jedním ze systémů se pak autorka přímo porovnává v kapitole 4 věnované evaluaci. K této kapitole bych ovšem měl výtku v tom smyslu, že kapitola spojuje dvě dosti nesouvisející evaluace: v první podkapitole jde o evaluaci vyvinutého nástroje Rhyme Tagger na anotovaných datasetech, zatímco druhá podkapitola evaluuje písňový korpus pomocí tohoto nástroje; jde tedy spíše o analýzu těchto textů než evaluaci a bylo by vhodnější ji oddělit to samostatné kapitoly.

Práce je spíše krátká, vlastní text má 41 stran.

Zdrojové kódy jsou poměrně přehledné a srozumitelné, opatřené komentáři a několika unit testy (konkrétně pěti). Chybí mi ovšem high-level přehled toho, co který skript dělá (k práci je prostě přiloženo 25 Pythonových skriptů v několika složkách).

Závěr

Původní zadání práce považuji za přehnaně ambiciózní, blížící se svou náročností spíše práci dizertační, a to mimo jiné z důvodu, že dané téma se zdá být dosud nedostatečně prozkoumáno. Z tohoto pohledu lze mít pochopení pro splnění pouze části tohoto zadání. Přesto se domnívám, že část zadání, která byla splněna, je poměrně malá, a spíše odpovídá kvalitní bakalářské práci, jako diplomovou práci ji tedy hodnotím spíše jako slabší. Autorka toto v práci zdůvodňuje tím, že mnoho času zabralo prozkoumávání uliček, které se ukázaly jako slepé.

Práce je zajímavá výzkumně a autorka zde navrhuje, implementuje a porovnává mnohé postupy. Zároveň jde o práci s nezanedbatelným implementačním rozsahem. Obě části jsou přitom provedeny poměrně kvalitně.

Lze tedy shrnout, že autorka splnila jen malou část zadání, ale zato ji splnila dobře.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 23. 8. 2021

Podpis: