

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Patricia Schmidtová		
Název práce	Generování divadelní hry		
Rok odevzdání	2022		
Studijní program	Informatika	Studijní obor	Matematická lingvistika
Autor posudku	Ondřej Dušek	Role	vedoucí
Pracoviště	Ústav formální a aplikované lingvistiky		

Text posudku:

Téma práce Diplomová práce Patricie Schmidtové se zabývá problémem automatického generování scénáře divadelní hry, tedy důležitým a složitým podproblémem v oblasti kreativního generování textu, resp. generování příběhů. Ta v posledních letech zažívá velmi rychlý rozvoj díky použití neuronových sítí, zejména architektury Transformer a předtrénovaných jazykových modelů na ní postavených. Většina dosavadních prací v oblasti se však věnuje mnohem kratším textům než je divadelní hra nebo i jen jedna scéna, navíc se specificky nezabývají generováním dialogů, ale cílí na příběhy v próze. Tyto úlohy dnes zdaleka nejsou vyřešeny – neuronové jazykové modely mají problémy zejména s udržení koherence příběhu. Velkým problémem zůstává také vyhodnocení vygenerovaných textů, kde selhávají automatické metody používané pro úlohy s pevnějším zadáním, jako je strojový překlad nebo generování textu z dat. Benchmarky pro kreativní generování jsou proto dnes velmi omezené a používají např. jen porovnávání návrhů místo plného generování.

V tomto ohledu je zadání této diplomové práce velmi ambiciózní – řeší nepříliš prozkoumaný problém a navíc jde daleko za běžnou délku vygenerovaného textu, se kterou dnešní systémy pracují. Navíc má práce i přímé napojení na výzkum a její výsledky jsou využity v praxi – byla vytvořena jako součást výzkumného projektu THEaiTRE, řešeného na ÚFAL, který tak autorce dodal určitou podporu (např. v otázce získávání trénovacích dat). Vzhledem k aplikaci v projektu a skutečnému uvedení vygenerované hry v divadle, ke kterému došlo v únoru 2021 a následně ještě v červnu 2022, práce počítá i s možností úprav a lidských zásahů do procesu generování, tj. kolaborací stroje a člověka.

Shrnutí obsahu Práce využívá aktuálních předtrénovaných jazykových modelů, zejména modelu GPT-2, a kombinuje je s několika typy vylepšení specificky zaměřených na generování divadelní hry:

- Použití speciálních promptů pro generování formátu dialogu a zároveň dotrénování modelu na doménově specifických datech (scénářích divadelních her a filmů), což umožňuje použít standardní model GPT-2 k „plochému“ generování.
- Nový hierarchický postup generování ve dvou krocích: První model GPT-2 dotrénovaný pro generování synopsí her vygeneruje kostru příběhu, druhý model na základě částí kostry generuje dialogy jednotlivých scén. Aby mohla autorka natrénovat generování dialogů podmíněných částmi kostry příběhu, musela navíc vytvořit i algoritmus pro propojení scén s relevantními částmi synopsí.
- Specifická vylepšení pro práci s postavami ve hře, motivovaná následujícími problémy: Jména postav jsou z hlediska jazykových modelů málo častá slova, která jednak mohou (více méně náhodně) ovlivnit postup generování a jednak není zaručena jejich konzistence napříč hrou. Autorka proto implementovala omezení dekódování modelu pro generování jména následující postavy v dialogu a zároveň upravila trénovací data pro anonymizaci jmen.
- Evaluace konzistence dialogu pomocí jazykové inference (NLI). Současné neuronové modely pro NLI jsou schopny s velkou přesností identifikovat kontradikce v textu, jejich aplikace v generování hry se proto nabízí. Autorka navrhla způsob, jak NLI použít jednak pro vyhodnocení generovaného textu a jednak přímo v generátoru pro vyloučení kontradikcí na výstupu.

Práce zahrnuje popis všech těchto vylepšení – včetně návaznosti na dosavadní publikace v oboru – a jejich experimentální vyhodnocení. Autorka se musela vypořádat s velikostí neuronových modelů a dlouhou dobou jejich trénování, s omezeným množstvím dostupných dat (synopsí a scénářů divadelních her a filmů) a problémy s jejich kvalitou, ale hlavně s omezenými možnostmi evaluace. Zde byly použity základní automatické textové metriky v kombinaci s lidským hodnocením za pomoci několika dobrovolníků.

Všechny problémy se podle mého názoru povedlo vyřešit v rámci možností výborně. Vyhodnocení by sice bylo dobré provést s větším vzorkem výstupů i hodnotitelů, ale je naprosto zjevné, že taková akce by logistikou i finanční náročností přesahovala rámec možností diplomové práce. Výsledky v každém případě posouvají hranice současného poznání. Některé implementované přístupy (hierarchické generování, dotrénování modelů) sice nezlepšují výsledek oproti použití většího předtrénovaného modelu, ale dovedou se mu v mnoha ohledech vyrovnat při výrazně nižší výpočetní náročnosti. Další implementovaná vylepšení (NLI, omezení postav) se jeví jako jednoznačně pozitivní. Autorka navíc podává dobrý popis limitů implementovaného modelu a navrhuje, jak je do budoucna překonat.

Struktura textu Práce obsahuje celkem 8 číslovaných kapitol. 1. kapitola podává úvod do práce i souvisejícího projektu THEaiTRE, kapitoly 2 a 3 představují potřebné teoretické pozadí a dosavadní relevantní publikace v oboru. Samotná implementace a experimenty jsou představeny v kapitolách 4-7. Nejprve se práce zabývá NLI (vzhledem k jeho použití pro evaluaci), pak základním úpravám „plochého“ generování, dále je představeno hierarchické generování včetně přípravy a propojení dat, a nakonec další rozšíření (omezení postav, použití NLI pro generování). Poslední 8. kapitola představuje krátké shrnutí a výhled do budoucna. Práce je psaná velmi dobře srozumitelnou angličtinou, struktura textu je logická a související oddíly na sebe vzájemně odkazují.

Průběh prací Autorka byla součástí týmu v projektu THEaiTRE od jeho počátku, práce na experimentech souvisejících s touto diplomovou prací tedy trvaly s různou intenzitou a některými odbočkami více než dva roky. Během experimentů autorka iniciativně hledala řešení problémů a konzultovala je se mnou i ostatními členy týmu projektu THEaiTRE. Samotný text práce vznikl letos na jaře, autorka na něm pracovala velmi intenzivně a detailně ho se mnou konzultovala. Všechny mé připomínky jsme společně vypořádali ještě před odevzdáním práce. Se spoluprací s autorkou jsem tedy naprosto spokojený.

Celkové hodnocení Práce zcela splnila zadání a převyšuje nároky běžně kladené na diplomovou práci. Jedná se o výzkumnou práci na nejvyšší úrovni, která posouvá současné poznání v oblasti kreativního generování textu. Na základě experimentů provedených v této diplomové práci vznikly důležité výstupy projektu THEaiTRE, mj. bylo možné inscenovat dvě divadelní hry vygenerované automaticky s minimálním množstvím lidských zásahů. Kromě toho se zde popsané experimenty staly základem několika publikací na mezinárodních konferencích, z nichž nejpodstatnější jsou tyto dva:

- R. Rosa, P. Schmidtová, A. Zachtarenko, O. Dušek, T. Musil, D. Mareček, S. Obaid, M. Nováková, K. Vosecká, D. Hrbek & D. Košťák, “THEaiTRobot: An Interactive Tool for Generating Theatre Play Scripts,” in *Proceedings of the 15th International Natural Language Generation Conference (INLG)*, Waterville, ME, USA, Jul. 2022.
- R. Rosa, P. Schmidtová, O. Dušek, T. Musil, D. Mareček, S. Obaid, M. Nováková, K. Vosecká, J. Doležal, “GPT-2-based Human-in-the-loop Theatre Play Script Generation,” in *Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*, Seattle, WA, USA, Jul. 2022, pp. 29–37.

Výsledky práce se tedy jistě stanou základem dalšího výzkumu v oboru. Plánujeme je využít i sami na pracovišti, kde připravujeme možné pokračování projektu THEaiTRE.

Celkově práci jednoznačně doporučuji k obhajobě; nemám žádné dotazy.

Práci doporučuji k obhajobě.

Práci navrhuji na zvláštní ocenění.

Tato diplomová práce si určitě zaslouží soutěžit o zvláštní ocenění. Jak už jsem napsal v hodnocení výše, přesahuje běžný rámec diplomové práce a přináší vědecké výstupy na nejvyšší úrovni, jak dokumentují i výše zmíněné publikace. Práce je nejen perfektní ukázkou samostatného výzkumu, ale i širší spolupráce na výzkumném projektu.

V Praze dne 26. 8. 2022

Podpis: