

Posudek vedoucího diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Bc. Vladislav Stankov

Název práce Speech-Informed Inverse Text Normalization

Rok odevzdání 2024

Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku doc. RNDr. Ondřej Bojar, Ph.D. **Role** vedoucí

Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Diplomová práce Vlada Stankova se zabývá částečně opomíjenou komponentou systémů pro automatické zpracování řeči: tzv. obrácenou normalizací (inverse text normalization, ITN), tj. komponentou která rozpoznanou posloupnost vyřčených slov upraví do podoby očekávané v psaném textu. I pokud se soustředíme jen na dvě nejnápadnější součásti této úlohy, doplnění interpunkce a kapitalizace písmen, je intuitivně jasné, že by zvuková stopa mluveného projevu měla být velmi relevantní. Běžně užívané metody ITN však zvuk zcela ignorují a predikují interpunkci a kapitalizaci jen na základě jazykového modelu, posloupnosti slov.

Vladova práce měla za úkol tento nedostatek odstranit a Vlad přitom experimentoval konkrétně s češtinou. Pro řešení zadaného úkolu Vlad použil současné moderní přístupy a vytvořil tak multimodální systém, který kombinuje bloky ze samotného rozpoznávání mluvené řeči (MFCC, konvoluce) s bloky jazykového modelování pomocí tranformerových enkodérů. Díky Vladovu pečlivému přístupu ve všech krocích práce se tak s jistotou dozvídáme, že zvuková stopa skutečně predikci interpunkce pomáhá. Tato pomoc je nejspolehlivěji patrná pro dostatečně četné typy interpunkce (především čárka a tečka).

Rád bych vyzdvihl zejména ucelenost Vladovy práce. ITN využívající zvukovou stopu vyžaduje pro supervizované učení specifická trénovací data, která přirozeně pro češtinu nebyla k dispozici. Konkrétně je třeba zvuk a doslovný přepis včetně správné interpunkce. Existující řečové korpusy však mají text rovnou korigován. (Jinými jazyky jsme se nezabývali, ale vzhledem k tomu, že tento druh ITN není příliš rozvíjen, je velmi pravděpodobné, že tato data nebudou často dostupná.) Vlad proto vytvořil novou verzi korpusu ParCzech 3.0 vhodnou pro tyto účely. Navazoval přitom na svou předchozí práci, byl autorem podoby ParCzechu pro trénování modelů rozpoznávání mluvené řeči (ASR) a významným způsobem korpus ve starší i teď nové verzi korpusu čistil.

Druhým přípravným krokem bylo vytvoření základního ASR, které by dostatečně kvalitně rozpoznalo vyřčená slova a především poskytlo přesnou časovou informaci. I tento obsáhlý krok

provedl Vlad s velkou pečlivostí.

Pro samotný ukol informovaného ITN pokládám Vladův přístup za inovativní. Kromě pozitivních výsledků je třeba vyzdvihnout velmi podrobnou analýzu chování modelů (např. vizualizace jejich attention) a též kontrastivní baseline. Úprava architektury neuronové sítě (přidání zvukové modality) dává totiž síti prostor využít více volných parametrů a dosáhnout oproti baseline lepších výsledků z jiných důvodů než díky dodatečné modalitě. Vladův experiment v kapitole 5.8 s pokaženým zvukem dokazuje, že to tak není.

Měl-li bych hledat největší slabinu, bude jí průběh vzniku textová podoby práce, nikoli však výsledký text. Text je kvalitní a bohatě doplněný ilustracemi, schémata a přehledně prezentovanými výsledky. Text vznikl poměrně pracně, hlavně jsme s Vladem společně obtížněji nacházeli formulace, které by byly dostatečně přesné a výstižné, a přitom ne příliš dlouhé. Je však třeba uznat, že práce se i svým textovým rozsahem řadí mezi diplomové práce obsažnější, a je proto přirozené, že její napsání trvalo déle.

Celkově Vlad pracoval na své diplomové práci dlouhodobě a svědomitě. Ačkoli to Vlad nezdůrazňuje, osobně pokládám i Vladův významný příspěvek ke korpusu ParCzech a příslušnému článku publikovanému na konferenci TSD za součást výsledků diplomové práce.

Výsledná diplomová práce jednoznačně prokazuje, že Vlad Stankov je schopen samostatné vědecké práce. Jeho příspěvek je hodnotný a nejméně aktualizovanou verzi korpusu ParCzech budeme publikovat jako řečový dataset. Práci jednoznačně doporučuji k přijetí.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 7. 6. 2024

Podpis: