

Oponentský posudek disertační práce

Západočeská univerzita v Plzni, Fakulta aplikovaných věd
Obor: Kybernetika
Student: Ing. Jan Lehečka
Název práce: Online topic-based language model adaptation

Oponent: doc. RNDr. Pavel Pecina, Ph.D.
Pracoviště: Ústav formální a aplikované lingvistiky, MFF UK, Praha

Obsah práce

Disertační práce Ing. Jana Lehečky se zabývá adaptací jazykových modelů v systémech rozpoznávání mluvené řeči s ohledem na téma. Myšlenka dynamicky měnit jazykový model v závislosti na vstupních datech není nijak nová a v literatuře lze najít příklady více či méně úspěšných pokusů, jak ji realizovat. Jan Lehečka se ve své práci zaměřil na použití v tzv. online systémech, tedy v situaci, kdy je řeč potřeba rozpoznávat průběžně s minimálním zpožděním, což klade na výsledný systém značná omezení. Jednak je potřeba rozpoznávat po krátkých úsecích, při rozpoznávání se není možné vracet a detekci témat a adaptaci jazykového modelu je tak potřeba provádět rychle. Předkládaná práce obsahuje několik originálních myšlenek a experimenty ukazují zlepšení kvality rozpoznávání ve srovnání s neadaptovaným modelem, a to na dvou úlohách: živý přepis televizního zpravodajství a živý přepis televizních sportovních přenosů, obojí v češtině. Práce obsahuje popis množství experimentů a jejich evaluace, včetně diskuse výsledků.

Struktura práce

Práce je vhodně strukturovaná, členěná do celkem 10 kapitol, opatřená seznamem citované literatury, bibliografií autora a technické přílohy popisující konfiguraci použitého systému pro rozpoznávání řeči. První kapitola obsahuje stručný úvod do studované problematiky. Druhá kapitola jasně specifikuje výzkumné cíle práce. Třetí kapitola poskytuje teoretický úvod do problematiky rozpoznávání mluvené řeči, identifikace témat a adaptace jazykových modelů. Čtvrtá kapitola obsahuje přehled existujících relevantních metod a publikovaných prací. Pátá kapitola popisuje navrhané řešení problému a jeho implementaci. Šestá kapitola je věnovaná experimentům s přepisem televizního zpravodajství a sedmá kapitola popisuje experimenty s přepisem televizních sportovních přenosů. V osmé kapitole jsou shrnuty celkové výsledky, v deváté kapitole je provedena jejich diskuse a desátá kapitola práci uzavírá shrnutím dosažených výsledků, nástinem možností pokračování výzkumu v této oblasti a autorovým vyjádřením k inovativnosti a praktičnosti odvedené práce a míře splnění jejích cílů.

Výsledky práce

Je potřeba říct, že cíle práce byly splněny, autorovi se podařilo navrhnout a implementovat řešení pro tématickou online adaptaci jazykového modelu v systému pro rozpoznávání mluvené řeči. Tento systém byl otestován, analyzován a prokázalo se, že zlepšuje kvalitu automatického přepisu ve dvou úlohách. K použitému řešení a výsledkům mám ale několik otázek a výhrad:

- 1) U dosaženého zlepšení byla sice prokázána statistická signifikance, může se ale snížení WER o méně než 1% (absolutně) s ohledem na vynaložené úsilí vyplatit v praxi?
- 2) Práce vychází z předpokladu, že obecný jazykový model použitý v systému pro rozpoznávání řeči je nedokonalý a v situaci, kdy se používají (tématicky) specifické výrazy, tak je nedokáže správně rozpoznat z důvodu podhodnocení odhadu jejich pravděpodobnosti (dané jazykovým modelem). V práci chybí analýza závažnosti tohoto problému, tedy do jaké míry je špatné rozpoznání slov zapříčiněno právě touto chybou jazykového modelu. Podobné analýzy se označují jako *oracle experiment* a obvykle stanovují horní hranici



potenciálního zlepšení. Lze odhadnout, jaký je prostor pro možné zlepšení v úlohách řešených v předkládané práci a do jaké míry se podařilo tento prostor využít?

3) Autor v práci používá klasické n-gramové jazykové modely, omezené na maximální délku $n=3$ a s omezením na minimální frekvenci n-gramů. Toto omezení autor vysvětluje potřebou mít jazykový model co nejmenší, protože jich potřebuje ukládat (kešovat) velké množství. Není ale jasné, jestli by n-gramové modely o vyšším řádu nefungovaly lépe a adaptace tak nebyla zbytečná. Podobně by bylo možné se ptát na použití neuronových modelů, které lépe pracují s delšími závislostmi, ale jejichž použití se autor vyhýbá.

4) Jedním z hlavních pilířů práce je tematická klasifikace textů, ale řešení této úlohy není nikde (intrinsicky) evaluováno. Jak dobře umí použitý klasifikátor klasifikovat dokumenty podle témat? Jak dobře umí přiřadit témata fragmentům rozpoznané řeči (ve srovnání např. s člověkem)? Je vůbec možné smysluplně přiřazovat takto krátké úseky přepsané řeči do témat použitých v experimentech?

5) Autor věnoval značnou pozornost paměťové a časové optimalizaci použitého řešení. Způsob vytváření a kešování interpolovaných modelů ale není příliš škálovatelný. Navíc samotné vytváření interpolovaných modelů a jejich výměna celý systém velmi omezují. Nebylo by vhodnější řešit interpolaci jazykových modelů skutečně online, tedy tak, že by všechny tematické modely a jeden obecný byly v paměti a při každém použití by byla provedena interpolace pravděpodobností z každého modelu s aktuálními interpolačními koeficienty? Odpadlo by náročné kešování a výměna modelů, interpolace by se mohla provádět bez omezení na hodnoty koeficientů. Obecný model by mohl být velký, tematické relativně malé.

6) V práci je na několika místech použita volba hyperparametrů celého systému ad-hoc, bez diskuse a analýzy možného efektu na výsledek (např. omezení na prvních 500 slov pro LSTM klasifikaci, uniformní rozdělení interpolační koeficientů tematických jazykových modelů, řád jazykového modelu, který už byl zmíněn, odstranění řídkých slov ze slovníku, parametry vážení slov v klasifikačním okně a další). Provedení chybějících analýz by přispělo ke zvýšení důvěryhodnosti použitého přístupu.

Jazyková a grafická úroveň

Práce je psaná anglicky, bez překlepů a hrubých gramatických chyb. Problémem je snad jen místy nesprávné používání určitých/neurčitých členů, které ale na porozumění textu nemá velký vliv. Jinak je text dobře strukturovaný, čitelný a srozumitelný.

Publikační činnost autora

Seznam publikací Jana Lehečky obsahuje 5 položek, z toho 4 příspěvky na konferenci TSD (ve 3 případech je Jan Lehečka uveden jako první autor), a jeden článek v impaktovaném časopise LREV, který považuji za hodnotný výsledek (i když Jan Lehečka není uveden jako první autor). Škoda, že alespoň některé z výše uvedených konferenčních příspěvků se nepodařilo prosadit na prestižnějších fórech v zahraničí.

Závěr

I přes uvedené nedostatky považuji předkládanou práci za přínosnou. Oceňuji zejména množství provedených experimentů a aplikování metod na dvě různé úlohy. Cíle práce byly splněny, výsledky publikovány na mezinárodních fórech a práci tedy doporučuji k obhajobě.

Pecina



**FAKULTA
INFORMATIKY**

Masarykova univerzita

Západočeská univerzita v Plzni

Doručeno: 04.06.2019

ZCU 014129/2019

listy: 4

druh:

přílohy:



zcupes1223a f2

POSUDEK OPONENTA DISERTAČNÍ PRÁCE

Název práce: *Adaptace jazykového modelu na téma v reálném čase*

Autor práce: *Ing. Jan Lehečka*

Oponent: *Doc. Mgr. Pavel Rychlý, Ph.D.*

Zhodnocení významu práce

Předkládaná práce rozvíjí obor automatického rozpoznávání řeči, kde již existují více-méně funkční řešení, je tam ale stále dost prostoru pro zlepšení. Práce popisuje jeden z možných směrů pro zlepšení a pro některé aspekty tohoto směru navrhuje dobrá řešení a ověřuje jejich skutečný pozitivní přínos.

Největší přínos práce vidím v provedeném vyhodnocení, vytvořené datové sady mohou být bezesporu použity při dalším vylepšování aplikací či systémů automatického rozpoznávání řeči.

Postup řešení a splnění cíle

Student v teoretické části práce přehledně popsal jednotlivé přístupy použitelné pro zadaný problém. Dále navrhl vlastní řešení pro zadané cíle. Navržený systém je vyhodnocen a výsledky přesvědčivě dokládají zlepšení oproti stávajícímu systému a tedy splnění cílů práce.

K práci mám následující dotazy/připomínky:

1. Jsou všechna témata (topics) z testovacích dat dostatečně homogenní? Je např. perplexita modelu pro jedno téma nižší než perplexita obecného modelu na celých datech? Jak se liší úspěšnost (WER, PNER) na jednotlivých tématech?
2. Praktické použití systému naráží na rychlost načtení jazykového modelu (sekce 5.6.5 a 6.5.6 na str. 74). Bylo by možné jazykový model vytvářet v podobě, v jaké s ním pracuje dekodér? To by umožnilo mapování dat do paměti (mmap v Unixu) bez nutnosti načítání celého modelu.
3. Všechny metody detekce tématu používají jiné struktury pro učení než jsou potom použity při detekci. Typicky jsou při učení použity celé příspěvky, klasifikátory se tedy učí na celém textu z daného příspěvku. Při rozpoznávání je ale z praktických důvodů použita jen relativně malá část z příspěvku. Pokud by trénování používalo také jen malou část, mohlo by být rozpoznávání úspěšnější.

Masarykova univerzita, Fakulta informatiky

Botanická 554/68a, 602 00 Brno, Česká republika

T: +420 549 49 1810, E: info@fi.muni.cz, www.fi.muni.cz

Bankovní spojení: KB Brno-město, ČÚ: 85636621/0100, IČ: 00216224, DIČ: CZ00216224

Výsledky práce a přínos předkladatele

Nejhodnotnějším přínosem disertační práce je vyhodnocení navržených metod. V rámci vyhodnocení byly vytvořeny dvě datové sady ze dvou různých domén, které pokrývají zaměření cílové aplikace. Výsledky vyhodnocení tedy budou věrně odpovídat úspěšnosti navržených metod při reálném použití v praxi. Samotné vyhodnocení je provedeno velice pečlivě, student změřil chování svého systému v řadě konfigurací a přehledně porovnává jejich úspěšnost. Kromě běžného *Word Error Rate* jsou měřeny i nově navržené *Proper Noun Error Rate* a *Name Error Rate*, které lépe vystihují zamýšlené použití systému. Druhým výrazným přínosem je způsob vytváření adaptovaného jazykového modelu na pozadí běžícího rozpoznávání. To umožňuje nasazení systému v reálném čase, na což předchozí systémy nebyly použitelné.

Popsaná zlepšení úspěšnosti navrženého systému oproti stávajícímu nejsou dramatická, nicméně jsou významná a jistě si zaslouží zavedení do praxe. Z tohoto pohledu je pozitivní přínos studenta k tématu automatického rozpoznávání řeči nesporný.

Formální a jazyková úroveň

Obsahově je práce dobře strukturovaná, s jasnými cíly i závěry. Text je vyvážený, v rozumném poměru obsahuje teoretické informace i technický popis konkrétních řešení či experimentů. Text je doplněn řadou ilustrací, jednotlivé grafy a obrázky pěkně ilustrují či dokládají diskutované skutečnosti.

Práce je psána velmi dobrou angličtinou, sazba textu je též kvalitní.

Publikace studenta

Některé části práce či příbuzná témata student publikoval na konferenci *Text, Speech and Dialog (TSD)* (celkem 4 články v letech 2011 až 2018) a v časopise *Language resources and evaluation* (v roce 2014). Tento časopis jistě tvoří kvalitní mezinárodní fórum, ale v něm studentem publikovaný článek není hlavním tématem předložené práce. Konference TSD je bezesporu mezinárodní, ale svým umístěním v České republice přitahuje z velké části domácí publikum. Další práce autora byly publikovány na vyloženě lokálních fórech. Celkově předložené publikace nejsou oslňující, nicméně dokládají solidní základ kvality předkládané práce.

Doporučení

Víše uvedené nedostatky nejsou zásadního charakteru, celkově je práce kvalitní. Doporučuji tedy disertační práci k obhajobě.

V Brně, 24. května 2019



Pavel Rychlý