

Oponentský posudek disertační práce

Západočeská univerzita v Plzni, Fakulta aplikovaných věd

Obor: Kybernetika

Student: Ing. Jan Švec

Název práce: Diskriminativní model pro porozumění mluvené řeči

Oponent: RNDr. Pavel Pecina, Ph.D.

Pracoviště: Ústav formální a aplikované lingvistiky, MFF UK, Praha

Obsah práce

Předkládaná práce Ing. Jana Švece se zabývá problematikou porozumění mluvené řeči (promluvy) v kontextu dialogových systémů. Autor se zaměřuje na několik vybraných oblastí výzkumu. Jednak jde o možnost zpracování neurčitostí, a to jak na výstupu (produkování více hypotéz významu promluvy), tak na vstupu (možností zpracování více hypotéz slovního přepisu promluvy). Dále se autor zabývá využitím rozpoznávání řeči na fonémové (nikoli slovní) úrovni k řešení nedostatečného pokrytí jazykového modelu, což bývá problém zejména pro morfologicky bohaté jazyky. Dále je to diskriminativní modelování porozumění mluvené řeči jako alternativa běžně používaným generativním modelům a kombinování statistických a pravidlových (znalostních) přístupů. Nedílnou součástí práce je popis experimentů a evaluace navržených metod a modelů.

Struktura práce

Práce je vhodně strukturovaná, členěná do 11 kapitol, opatřená seznamem citované literatury a bibliografií autora. První kapitola obsahuje úvod práce a její zasazení do kontextu dialogových systémů. Jejich detailnější popis je uveden v kapitole druhé. Kapitulu třetí tvoří přehled současného stavu řešené problematiky a kapitola čtvrtá jasně a přehledně stanovuje cíle předkládané disertační práce (celkem v pěti bodech). Kapitola pátá pokládá teoretický základ použitých metod. Vlastní přínos práce je potom tvořen kapitolami šest až deset, které popisují navrhovaný diskriminativní model porozumění mluvené řeči (kapitoly 6 a 7), model detekce sémantických entit (kapitola 8) a experimenty ověřující použitelnost obou metod a jejich kombinace (kapitoly 9 a 10). Závěr práce tvoří kapitola jedenáctá, ve které se autor vyjadřuje k míře splnění jednotlivých cílů disertační práce a možným směrům dalšího výzkumu.

Výsledky práce

Myšlenka použití diskriminativních modelů pro porozumění mluvené řeči není nová. Autor ji ale obohacuje o tzv. konceptový model, který vstupní promluvě přiřazuje abstraktní sémantický strom, který je (částečně) zarovnán s lexikální realizací detekovaných sémantických entit, a to v jediném diskriminativním modelu. Hierarchický diskriminativní model (HDM) navržený autorem rozšiřuje model STC (klasifikátor sémantických n-tic) o možnost zpracování neurčitého vstupu (prezentovaného formou mřížky) a predikce více výstupních hypotéz (sémantických stromů). Metoda detekce sémantických entit potom kombinuje znalostní (expertní) a statistický přístup. Navržené modely a metody byly experimentálně ověřeny a na

dvou úlohách: podávání informací o odjezdech a příjezdech vlaků a inteligentní telefonní asistentka a téměř ve všech případech vedly ke statisticky signifikantnímu zlepšení výsledků. Kromě celkové úrovně práce lze vyzvednout vynikající zpracování teoretických částí (jak v úvodu práce, tak při popisu experimentů) a také kvalitu provedených experimentů a detailní analýzu výsledků. Za významné považuji i dílčí výsledky, které měly vliv na ty celkové (např. adaptace fonémových jazykových modelů, výpočet racionálních jádrových funkcí, nastavení parametrů HDM). Pokud bych měl práci něco vytknout, tak je to fakt, že experimenty byly provedené pouze na češtině. Bylo by zajímavé ověřit tyto výsledky i na jiných jazycích, zejména např. použití fonémových modelů.

Jazyková a grafická úroveň

Práce je psaná česky, obsahuje jen minimum pravopisných chyb a překlepů. Text je dobře čitelný a srozumitelný. Grafická úroveň práce je zcela v pořádku.

Publikační činnost autora

Seznam publikací Jana Švece obsahuje výrazně nadprůměrných 24 položek, včetně příspěvků na velmi prestižních konferencích (např. Interspeech a ICASSP) a článku v odborném časopise. Prvním autorem je Jan Švec přibližně u poloviny publikací.

Závěr

Předkládaná disertační práce je významným příspěvkem v oblasti porozumění mluvené řeči a dialogových systémů. Všechny její cíle byly splněny a publikovány na mezinárodních fórech. Práci jednoznačně doporučuji k obhajobě.



RNDr. Pavel Pecina, Ph.D.

15.11. 2013, Praha



Západočeská Univerzita v Plzni,
Fakulta aplikovaných věd
Obor: „Kybernetika“

Posudek na disertační práci

Ing. Jan Švec: Diskriminativní model pro porozumění mluvené řeči

A) *Aktuálnost zvoleného tématu*

Počítačový systém, který je schopen vést smysluplný dialog s člověkem, lze považovat za jednu z těch typických met umělé inteligence, o které se vědci pokoušejí víc než 50 let a jejichž dosažení se zdá být tím obtížnější, čím hlouběji vědci do dané problematiky pronikají. Domovské pracoviště Ing. Ševce patří k evropské špičce v oboru zpracování mluvené řeči. Díky tomu měl autor příležitost se podrobně seznámit s aktuálními postupy využívanými ve světě, otestovat je na reálných úlohách a identifikovat jejich přednosti i slabé stránky. První tři kapitoly disertace Ing. Švece na stranách 1-20 nabízejí stručný přehled existujících přístupů k porozumění mluvené řeči, který se opírá o úctyhodnou literární rešerši zmiňující 70 publikací i o vlastní experimenty realizované v letech 2007-2013 (např. publikace [53] a [58]). Tato zkušenost zřetelně ovlivnila formulaci tématu disertační práce, která se soustředí na hledání postupů pro zvýšení robustnosti současných dialogových systémů prostřednictvím vývoje efektivních metod, které umožní lépe porozumět mluvené řeči vztahující se k relativně úzce vymezené doméně.

Příležitost ke zdokonalení vidí Ing. Švec v tom, že je třeba se soustředit na vývoj postupu, který bude zohledňovat současně jak neurčitost při rozpoznání řeči, tak více možností pro přepis studovaného zvukového vstupu do slov a pak i do více významů. Takový přístup hrozí kombinatorickou explozí při zpracování konkrétních dat. Proto se Ing. Švec ve své práci soustředí na vývoj pravděpodobnostního diskriminativního modelu uvažovaných promluv, pro který se snaží omezit požadavky na rozsah trénovacích dat vtipným využitím znalostí o prostředí úlohy. Toto spojení znalostního a statistického přístupu je nepochybně zajímavé a v daném kontextu zcela původní. Téma celé práce je nejen aktuální, ale i velmi náročné, neboť vyžaduje jak hluboké porozumění řadě různorodých technik, tak důkladné experimentální ověření vlastností navržených řešení.

B) *Splnění cílů disertace*

Disertace si klade pět cílů, které jsou jasně vytyčeny v Kapitole 4 na stranách 21 a 22:

1. Vyvinutí modelu porozumění schopného pracovat s neurčitostí vstupu i výstupu.
2. Využití fonémového rozpoznávače v oblasti porozumění řeči.
3. Formulace plně pravděpodobnostního diskriminativního modelu.
4. Návrh vhodné metody pro kombinaci statistického a znalostního přístupu.
5. Ověření modelu nad více cílovými doménami.

Kapitola 5 (str. 23-56) vysvětluje teoretické principy algoritmů, které autor později používá jako součást vlastního řešení stanovených cílů. Speciální pozornost je věnována původní implementaci klasifikátorů sémantických n -tic, která je připravena pro práci s češtinou a testována při experimentech, které slouží jako motivace pro další vlastní práci autora. Jednotlivé kroky této práce jsou popsány a zhodnoceny v kapitolách 6-10 (str. 57-150).

Splnění cílů disertace je důkladně dokumentováno v závěrečné 11. kapitole na stranách 154 až 157, kde je ke každému z výše uvedených cílů uveden stručný postup řešení a případně i odkazy na velmi přesvědčivé experimentální výsledky uvedené v předchozích kapitolách. Cílů práce bylo nepochybně dosaženo.

C) Zvolené metody zpracování

Disertace vychází z důvěrné znalosti aktuálně používaných metod zpracování řečového signálu, které tvůrčím způsobem upravuje a doplňuje. Presentované výsledky odpovídají velkému časovému nasazení autora. Možná proto jsou některé formulace poněkud zkratkovité a některé části textu se mohou zdát čtenáři dost těžko přístupné. Nemohu nechat bez povšimnutí některé nedostatky čistě formálního charakteru, např.

- Nepovažuji za šťastnou formulaci na str. 25 před formulí (5.20): „... *případ, kdy nadrovina H není lineární.*“
- Při definici operace $*$ (ve smyslu Kleeneho uzávěru) na str. 47 by neměl chybět požadavek na to, že se jedná o konečné řetězce. v
- Ve formulaci podmínky na sumu za rovnicí (5.88) na str. 48 by měla být explicitně uvedena množina možných hodnot β .
- V rovnici (5.95) na str. 51 chybí vysvětlení významu pro použitý symbol α .
- Čtenáři by pomohlo, kdyby při hodnocení experimentů s různými modely byly v textu explicitně uvedeny odkazy na kapitoly, kde se o nich podrobněji píše. Např. v odstavci 10.6 (str. 149-150) by mělo být připomenuto, v čem je rozdíl mezi modely „slovní přepis“ a „slovní přepis (k)“, pro které se uvádějí křivky učení.

Autor navrhuje řadu zajímavých nových postupů. Byla bych ráda, kdyby odpověděl na následující otázky:

- Jak doporučuje postupovat při stanovení ceny editačních operací, kterou potřebuje pro definici editační vzdálenosti v kapitole 9?
- Pro volbu nejvhodnějších hodnot parametrů „váha jazykového modelu“ a „penalta vložení slov“ pro systém navrhovaný v části 9.4.6 je použito exhaustivní prohledávání, které je nepochybně poměrně náročné. Je možné pro volbu těchto hodnot nějaký jiný postup, třeba jen modifikovat hodnoty z podobných úloh?

D) Výsledky disertace a nové poznatky, které přináší.

Předložená práce otvírá několik slibných a zcela původních směrů dalšího výzkumu. Autorovi se podařilo experimentálně doložit, že navržený a vytvořený systém fonémového rozpoznávání je výrazně odolnější vůči chybám rozpoznávání akustického

signálu než běžně užívané postupy. To je nepochybně velmi atraktivní z hlediska praktických aplikací. Za velmi cennou vlastnost diskriminativního modelu Ing. Švece považuji to, jak přirozeně se mu do tohoto modelu daří integrovat různé doménové znalosti i to, že modely vytvořené pro jednu doménu lze poměrně lehce využít i v doménách dalších.

E) Význam výsledků pro praxi nebo další rozvoj vědy

Téma předložené disertace vycházelo z praktických zkušeností autora, který se běžně v praktických úlohách potýkal s tím, jak náročné je dosáhnout dostatečné robustnosti u současných dialogových systémů. A právě pro některé problémy související s tímto tématem práce nabízí elegantní a inovativní řešení. Například fonémový přístup přispívá ke zvyšování odolnosti těchto systémů vůči chybám při rozpoznávání zvukových signálů. Kombinace znalostního a statistického přístupu zase výrazně snižuje nároky na přípravu dat pro získání konkrétních hodnot parametrů použitého modelu. Je tedy zřejmé, že dosažené výsledky mají význam pro praxi.

F) Závěrečné hodnocení práce

Disertace splňuje podmínky samostatné tvůrčí vědecké práce a obsahuje původní autorem publikované výsledky uveřejněné na 21 mezinárodních konferencích (10 z nich jsou typu IEEE) a v 1 časopiseckém článku. Práci doporučuji k obhajobě.

V Praze, 13.3. 2014



Olga Štěpánková