Rozšířený Abstrakt

Tvorba systému rozpoznávání řeči pro angličtinu

Bc. Lukáš Matějů, Ing. Petr Červa, Ph.D.

Abstrakt

Diplomová práce se zabývá tvorbou systému rozpoznávání řeči pro angličtinu z hlediska akustického a jazykového modelování. Vysvětluje základní terminologii a přístupy k modelování. V rámci práce jsou shromážděna akustická a textová data pro angličtinu. Pomocí těchto dat a navržených skriptů jsou natrénovány modely, jejichž kvalita je experimentálně otestována. Jednotlivé experimenty reprezentují provedené úpravy za účelem zvýšení úspěšnosti rozpoznávání. Ta se v průběhu práce zlepšila o více jak 25 %. Nejlepší modely jsou převedeny do formátu aplikace Newton Dictate, jejíž základem je rozpoznávač vyvíjený na Technické Univerzitě v Liberci, kde by měly sloužit jako ukázka možností libereckého rozpoznávače.

Úvod

Laboratoř počítačového zpracování řeči Technické Univerzity v Liberci se zabývá tvorbou systému rozpoznávání řeči již od devadesátých let minulého století. Rozpoznávací systém vyvíjený v Liberci má pro český jazyk 95-97% úspěšnost u diktování libovolnou osobou při velikosti slovníku 350 000 slov. Hlavní motivací této práce je rozšíření libereckého rozpoznávače o další jazyk pro rozpoznávání, konkrétně o americkou angličtinu.

Rešeršní část práce by měla představit základy trénování akustických a jazykových modelů a vyhodnocování úspěšnosti rozpoznávání řeči. Pro angličtinu by následně mělo být shromážděno co největší množství akustických, jazykových a lexikálních dat. Pomocí vhodných nástrojů (toolkity HTK, Kaldi) by měly být navrženy trénovací skripty pro tvorbu akustických a jazykových modelů, tedy základních stavebních kamenů pro konkrétní jazyk. Vytvořené modely by následně měly být experimentálně srovnány a nejlepší z nich převedeny do prostředí Newton Dictate.

Experiment a metody

Veškeré experimenty byly prováděny pomocí libereckého rozpoznávače a rozpoznávače obsaženého v toolkitu Kaldi. Jednotlivé pokusy odráží provedené úpravy na akustickém i jazykovém modelu a jejich výsledky ilustrují vliv těchto úprav na úspěšnost rozpoznávání. Pro rozpoznávání byl použit slovník se 130 000 záznamy, který byl připraven pro tuto práci na základě 39 znakové fonetické abecedy popisující americkou angličtinu.

Vytvářené akustické modely pomocí skriptů a toolkitů HTK a Kaldi byly fonémové skryté markovské modely. Jako zdrojová data byly použity nahrávky z korpusů TIMIT a VoxForge, ke kterým byly získány fonetické přepisy s anotovanými hluky metodou pevného zarovnání. To také sloužilo k profiltrování nekvalitních záznamů, které nebyly vhodné k trénování.

Tabulka 1. Akustická data

korpus	trénovací data	testovací data
TIMIT	4,1 hod	1,3 hod
VoxForge	69,3 hod	2,8 hod

Rozšířený Abstrakt

Pro jazykové modely se staly zdrojem texty z informačních serverů Reuters a The Guardian, ze kterých bylo navrženým webovým pavoukem postahováno dostatečné množství textových dat. Tato data byla následně předzpracována metodami závislými i nezávislými na konkrétním jazyku. Následně byly natrénovány bigramové jazykové modely.

Tabulka 2. Jazyková data

MB	Reuters	The Guardian
data	1 560 MB	1 800 MB

Výsledky a diskuze

Následující tabulka ukazuje rozdíl v úspěšnosti rozpoznávání prvních vytvořených modelů založených jen na korpusu TIMIT s finálními modely, jejichž zdrojová data byla rozšířena a prošla výše popsanými úpravami.

Tabulka 3. Výsledky experimentu založeného na prvotních a finálních modelech

Accuracy	první modely	finální modely
TUL	38,99 %	65,76 %
Kaldi	40,18 %	66,32 %

Největší podíl na zvýšení úspěšnosti rozpoznávání měla anotace ticha a hluků a zavedení kvalitního jazykového modelu. Během práce se podařilo zlepšit úspěšnost rozpoznávání až o 25 % na stejné testovací sadě s mnohonásobně větším slovníkem. Úspěšnost rozpoznávání je u rozpoznávače vyvíjeného na Technické Univerzitě v Liberci a rozpoznávače přítomného v Kaldi velmi podobná a ověřuje tak správnost postupů vytváření modelů.

Závěr

V rámci diplomové práce byla shromážděna akustická, jazyková a lexikální data pro anglický jazyk. Tato data byla na základě vytvořených skriptů předzpracována a připravena pro použití při trénování a testování. Pro tvorbu modelů založených na angličtině byly navrženy trénovací skripty, které umožňují jejich snadné vytváření. Akustické i jazykové modely byly následně experimentálně otestovány. Úspěšnost nejlepších modelů při rozpoznávání přesahuje 65 %. Nejlepší modely byly převedeny do formátu použitelného v aplikaci Newton Dictate, zároveň je možné je použít pro zisk fonetických přepisů dalších akustických dat pro případné další rozšíření.

Na diplomovou práci by bylo možné dále navázat. Rozšiřování zdrojových akustický, lexikálních i jazykových dat by vedlo k dosažení ještě lepších výsledků úspěšnosti rozpoznávání. V úvahu také připadají alternativní méně používané metody vytváření akustických i jazykových modelů a následné porovnávání výsledků experimentů s výsledky prezentovanými v této práci.

Reference

[1] HUANG, Xuedong. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Vyd. 1. New Jersey: Prentice-Hall, 2001. ISBN 01-302-2616-5.

[2] *Řeč a počítač: sborník článků.* Vyd. 1. Editor Jan Nouza, Zbyněk Koldovský, Robert Vích. Liberec: Technická univerzita v Liberci, 2009, 235 s. ISBN 978-80-7372-548-8.

[3] *HTK Speech Recognition Kit* [online]. [2009] [cit. 2014-03-31]. Dostupné z: http://htk.eng.cam.ac.uk/