Studentská Konference Fakulty Mechatroniky, informatiky a mezioborových studií
27. květen 2014, Liberec, Česká republika

Tvorba systému rozpoznávání řeči pro angličtinu

TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií

Bc. Lukáš Matějů
Ing. Petr Červa, Ph.D.
Ústav informačních technologií a elektroniky

Abstract

This diploma thesis is dedicated to building a speech recognition system based on acoustic and language modelling. Basic terminology and approaches to modelling are explained. In the practical part of this paper acoustic and text data are gathered. Using these data and designed scripts acoustic and language models are trained. Their quality is experimentally tested. All of the experiments reflects individual changes made to models and source data in order to achieve higher recognition accuracy. The highest achieved accuracy is close to 66 %. The best experimentally chosen acoustic and language models are converted into file format supported by Newton Dictate application, which is based on recognition system developed at Technical University in Liberec.

Úvod

Zpracování řeči je progresivním oborem, který se stále více promítá do běžného lidského života. Pod tento obor patří i úloha rozpoznávání řeči, při které je převáděna lidská řeč do textové podoby. 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.

Cíle

Cílem teoretické části diplomové práce je seznámení s principy akustického a jazykového modelování a s metodami vyhodnocování úspěšnosti rozpoznávání. Hlavní cílem praktické části je shromáždění akustických, jazykových a lexikálních dat, vytvoření trénovacích skriptů a následné natrénování akustických a jazykových modelů, tedy základních stavebních kamenů pro konkrétní jazyk, pro spojité rozpoznávání angličtiny v prostředí Newton Dictate.

Metodika

Pro popis zvukové stránky anglického jazyka byla vybrána fonetická abeceda skládající se z 39 fonémů. Na jejím základě byl sestaven slovník pokrývající velkou část americké angličtiny. Zdrojem akustických dat je korpus TIMIT obsahující několik hodin nahrávek angličtiny. Pro vytvoření robustnějších modelů byla tato data rozšířena o nahrávky z projektu VoxForge (70 hodin záznamů). Shromážděná data byla pomocí skriptů parametrizována a podmnožina TIMIT byla použita k natrénování akustického modelu toolkitem HTK. Tento model byl použit v úloze pevného zarovnání pro zisk fonetických přepisů pro nahrávky VoxForge a pro anotaci ticha a různých hluků. Výsledky pevného zarovnání také posloužily k strojovému vyfiltrování nekvalitních akustických záznamů. Pro angličtinu byly navrženy trénovací skripty pomocí toolkitů HTK a Kaldi. Vytvářené modely byly fonémové skryté markovské modely.
Pro tvorbu jazykových modelů nebyl k dispozici žádný textový korpus. Z tohoto důvodu byl navržen webový pavouk, jehož cílem byl sběr textů z vybraných zpravodajských serverů. Takto získané texty (téměř 4 GB) byly následně automaticky předzpracovány. Výsledné trénovací skripty vytváří bigramové modely použitelné jak v rozpoznávači vyvíjeném na Technické Univerzitě v Liberci, tak v Kaldi.

Výsledky

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.
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 1. 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.

Závěry

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] Řeč a počítač: principy hlasové komunikace, úlohy, metody a aplikace. Vyd. 1. Editor Jan Nouza, Zbyněk Koldovský, Robert Vích. Liberec: Technická univerzita v Liberci, 2009, 235 s. ISBN 978-80-7372-548-8.
[2] JURAFSKY, Dan a James H MARTIN. Speech and Language Processing. 2nd ed. Upper Saddle River: Pearson Education, 2008, 988 s. ISBN 978-0-13-187321-6.
[3] Kaldi [online]. [2014] [cit. 2014-03-31]. Dostupné z: http://htk.eng.cam.ac.uk/

Kontakt

Bc. Lukáš Matějů, lukas.mateju@tul.cz

Tato práce byla podpořena z projektu
Studentské grantové soutěže (SGS)
na Technické univerzitě v Liberci v roce 2014