TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies

A Modern Look on Speech Processing

Thesis Statement

Study programme: P2612 – Electrical Engineering and Informatics
Study branch: 2612V045 – Technical Cybernetics

Author: **Ing. Lukáš Matějů**
Supervisor: Ing. Petr Červa, Ph.D.

Liberec 2017

Abstract

Abstract in English...

Keywords: A, B, C, D, E

Contents

List of Abbreviations

CNN	Convolutional Neural Networks
DNN	Deep Neural Networks
FAR	False Alarm Rate
FER	Frame Error Rate
HTER	Half-Total Error Rate
LID	Language Identification
MR	Miss Rate
SAD	Speech Activity Detection
SID	Speaker Identification
RNN	Recurrent Neural Networks
RTF	Real-Time Factor
VAD	Voice Activity Detection
WER	Word Error Rate
WFST	Weighted Finite State Transducers

Introduction

Úvodní kapitola seznamující s obsahem tématu.

- preprocessing řeči,
- detekce řeči/neřeči,
- detekce jazyka,
- změna mluvčího
- id mluvčího
- další segmentační úlohy, emoce
- -> rozpoznávání řeči s využitím získaných informací

Seznámení s rozvržením práce a současným stavem:

- členění práce - jednotlivé kapitoly obsahují následující body
- SAD - hotovo,
- LID - rozpracováno,
- D+SID - další v pořadí,
- případné další úlohy (emoce atd.).

Podobné systémy:

- na čem založené,
- v čem se liší,
- proč znovu?
- porovnání.
- LIAM, Alisé

1 An Overview of Field Research

Rešerše pro jednotlivé bloky. Od vybraných původních metod po state-of-the-art. Otázka... jak podrobně? - podle rozsahu ostatních kapitol?

1.1 SAD

state-of-the-art může mírně vycházet z úvodů článků, rozšíření + arpa challenges

1.2 LID

doba před ivectory, ivectory

1.3 D+SID

2 Evaluation Metrics

FER, MR, FAR, HTER, F-value, delta 2/3, RTF, rovnice

3 Proposed Solution

- představit na čem je založené...
- supervised machine learning, deep learning (torch, gpus), liší podle modulu
- podrobnější představení technologií? viz DNN apod.?

3.1 Employed Technologies

fbc, DNN, WFST (dekodér) (SAD)
fbc + jak bottleneck příznaky, ivectory a jak kombinovat s DNN (LID)
otázka? ... jak podrobně

3.1.1 Deep Neural Networks

3.1.2 Weighted Finite State Transducers

3.1.3 iVectors

3.2 Experimental Setup

styl použití technologií, idea, DNN na trénování, WFST vyhlazení atd.
základní nastavení, torch, náš rozpoznávač atd.

4 Development of the Proposed Solution

vývoj jednotlivých bloků, hlavní část tvoří SAD, zbytek LID.

4.1 SAD

- vychází z publikovaných článků. SIGMAP, ICASSP a SPRINGER,
- od začátku návrhu až po finální modely WFST. Vyhodnocení na QUT-NOISE-TIMIT, rozšířené testy ze springeru,
- vliv SADu na rozpoznávač,
- +/- vše publikované + pár dílčích věcí, co se nepoužilo.
- zdůraznění výsledků + že se používá na tul

4.2 LID

- bez ivectorů, fbc přínaky, bottleneck příznaky,
- ivectory úvod,
- (zatím nepublikováno)

Možno 'vykrást' publikované články.

5 Conclusions

Závěrečná kapitola shrnující celkový stav práce:

- SAD (hotovo) + zhodnocení výsledků [1] [2] [3],
- LID (rozpracováno) + zhodnocení dosavadních výsledků + co dál,
- D+SID (bude) + co dál,
- případné další úlohy.

Případné další směrování práce.

References

[1] L. Mateju, P. Cerva, and J. Zdansky, "Study on the Use of Deep Neural Networks for Speech Activity Detection in Broadcast Recordings," in *Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (ICETE 2016) - Volume 5: SIGMAP, Lisbon, Portugal, July 26-28, 2016*, pp. 45–51, 2016.

[2] L. Mateju, P. Cerva, J. Zdansky, and J. Malek, "Speech Activity Detection in online broadcast transcription using Deep Neural Networks and Weighted Finite State Transducers," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2017, New Orleans, LA, USA, March 5-9, 2017*, pp. 5460–5464, 2017.

[3] L. Mateju, P. Cerva, and J. Zdansky, "Investigation into the Use of WFSTs and DNNs for Speech Activity Detection in Broadcast Data Transcription," in *E-Business and Telecommunications - 13th International Joint Conference, ICETE 2016, Lisbon, Portugal, July 26-28, 2016, Revised Selected Papers*, pp. 1–18, 2017.

Author's Publications

2017:

1. L. Mateju, P. Cerva, and J. Zdansky, "Investigation into the Use of WFSTs and DNNs for Speech Activity Detection in Broadcast Data Transcription," in *E-Business and Telecommunications - 13th International Joint Conference, ICETE 2016, Lisbon, Portugal, July 26-28, 2016, Revised Selected Papers*, pp. 1–18, 2017.

2. R. Safarik and L. Mateju, "The Impact of Inaccurate Phonetic Annotations on Speech Recognition Performance," in *Text, Speech, and Dialogue - 20th International Conference, TSD 2017, Prague, Czech Republic, August 27-31, 2017, Proceedings*, pp. 1–8, 2017.

3. L. Mateju, P. Cerva, J. Zdansky, and J. Malek, "Speech Activity Detection in Online Broadcast Transcription Using Deep Neural Networks and Weighted Finite State Transducers," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2017, New Orleans, LA, USA, March 5-9, 2017*, pp. 5460–5464, 2017.

2016:

4. M. Bohac, L. Mateju, M. Rott, and R. Safarik, "Automatic Syllabification and Syllable Timing of Automatically Recognized Speech - for Czech," in *Text, Speech, and Dialogue - 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings*, pp. 540–547, 2016.

5. L. Mateju, P. Cerva, and J. Zdansky, "Study on the Use of Deep Neural Networks for Speech Activity Detection in Broadcast Recordings," in *Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (ICETE 2016) - Volume 5: SIGMAP, Lisbon, Portugal, July 26-28, 2016*, pp. 45–51, 2016.

6. R. Safarik and L. Mateju, "Impact of Phonetic Annotation Precision on Automatic Speech Recognition Systems," in *39th International Conference on Telecommunications and Signal Processing, TSP 2016, Vienna, Austria, June 27-29, 2016*, pp. 311–314, 2016.

2015:

7. L. Mateju, P. Cerva, and J. Zdansky, "Investigation into the Use of Deep Neural Networks for LVCSR of Czech," in *2015 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics, ECMSM, 2015, Liberec, Czech Republic, June 22-24, 2015*, pp. 184–187, 2015.