Aleksander Lempinen

Automatic speech recognition in physics teacher education

Master's Thesis in Information Technology

September 13, 2019

University of Jyväskylä

Faculty of Information Technology

Author: Aleksander Lempinen

Contact information: aleksander.lempinen@outlook.com

Supervisors: Tommi Kärkkäinen, Daniela Caballero, and Jouni Viiri

Title: Automatic speech recognition in physics teacher education

Työn nimi: Automaattinen puheentunnistus fysiikan opettajakoulutuksessa

Project: Master's Thesis

Study line: Educational Technology

Page count: 13+0

Abstract: TODO: Abstract

Keywords: TODO: Keywords

Suomenkielinen tiivistelmä: TODO: Tiivistelmä

Avainsanat: TODO: Avainsanat

Glossary

TODO	TODO: Glossary

Contents

1 Introduction

Modern physics classroom instruction attempts to improve learning by reducing the cognitive load, which is done by providing a clear organizational structure for the factual knowledge and linking new material to previously known ideas (Wieman and Perkins 2005). The goal of physics instruction is to help students become experts capable of solving problems (Fischer et al. 2014; Wieman and Perkins 2005). While lecture based instruction is often not very effective for retaining new knowledge (Wieman and Perkins 2005), the content structure and relationships between concepts presented by the teacher positively correlate with student learning gains when analysed with conceptual network analysis (Fischer et al. 2014).

For a long time automatic speech recognition (ASR) was outperformed by human speech recognition (HSR), but with the recent developments in the deep learning approach the error rate of automatic speech recognition and human speech recognition is almost the same in certain tasks (Spille, Kollmeier, and Meyer 2018). This however is language specific. Work on automatic speech recognition with conversational Finnish started in 2012 and word error rate of 27.1% was achieved by 2017 (Enarvi 2018).

In natural language processing the creation of treebanks such as Turku Dependency Treebank and FinnTreeBank in the past decade have allowed for development of natural language processing toolkits with adequate performance with Finnish language such as FinnPos (Silfverberg et al. 2016) and TurkuNLP (Kanerva et al. 2018). Lemmatization in the new toolkits is of particular interest, because it is essential for real world tasks in inflective languages such as Finnish (Kanerva et al. 2018).

Manually transcribing and preprocessing lessons from audio data for analysis is a very laborious task. The aim of this study is to develop a pipeline for studying physics lesson content structures and relationships between physics concepts with conceptual network analysis using automatic speech recognition and natural language processing. This is done to provide a new tool for analysing physics instruction for research and teacher education purposes.

2 Physics instruction quality

2.1 Pedagogical link making

2.2 Conceptual network analysis

3 Speech as data

3.1 Automatic speech recognition

Automatic speech recognition (ASR), sometimes called speech to text, is a classification task, where the goal is to predict what was said from the audio signal of speech. Early ASR systems had an acoustic model which detected different sounds also known as phonemes to recognize numbers, some vowels and consonants for a single speaker (Juang and Rabiner 2005). Improvements in the acoustic model allowed for introduction of speaker-independent ASR (Benzeghiba et al. 2007; Juang and Rabiner 2005). The later addition of a language model based on statistical grammar and syntax helped more accurately predict the correct word based on what words previously appeared in the sentence (Juang and Rabiner 2005). Modern ASR systems utilize the fact that sentences are sequences of words and words are sequences of phonemes (Bengio and Heigold 2014).

Sequence based models such as hidden markov models (HMM) are most commonly used, but deep learning approaches using recurrent neural networks (RNN) and long short-term memory (LSTM) networks gaining popularity for acoustic models, language models and end to end text-to-speech models (Bengio and Heigold 2014; Enarvi 2018). Deep learning approach is more data-driven and relies on fewer assumptions, but instead requires more data for training (Bengio and Heigold 2014). This might be impractical if training data is limited. Depending on the architecture, an ASR system might be capable of either transcription, keyword spotting or both (Juang and Rabiner 2005; Enarvi 2018).

ASR is a difficult machine learning task because of a large search space, large vocabulary, undetermined length of word sequences and problems related to aligning speech signal to the text (Enarvi 2018). Speech is highly variable even with a single speaker due to noise, but different pronounciations and accents mean that the audio signal will be different despite the same words being spoken (Juang and Rabiner 2005). Accents, dialects, emotional state, gender and casual speech slurring in spontaneous speech bring a lot of variation which makes conversational speech especially difficult compared to standard pronounciations and vocabulary (Benzeghiba et al. 2007; Juang and Rabiner 2005). Speaker-dependent systems are

typically more accurate than speaker-independent systems (Benzeghiba et al. 2007; Enarvi 2018).

3.2 Natural language processing

TODO: NLP overview (Silfverberg et al. 2016; Kanerva et al. 2018)

3.3 Finnish speech

TODO: Conversational vs official Finnish Finnish language is particularly difficult for ASR because words are formed by concatenating smaller (**enarvi**)

4 Methods

4.1 Data

The data set consists of 25 Finnish physics lessons on the topic of "Relation between electrical energy and power" from Quality of Instruction in Physics (QuIP) project (Helaakoski and Viiri 2014). In addition to the teacher speech recording, test results from each student are available from before and after the lesson.

4.2 Preprocessing

4.2.1 Network analysis

4.2.2 Adjacency matrix

4.3 Supervised machine learning

4.4 Clustering

4.5 Validation

5 Results

TODO:

6 Conclusions

TODO:

Research is being done in collaboration with the Department of Teacher Education at University of Jyväskylä and Centro de Investigación Avanzada en Educación (CIAE) at University of Chile.

Bibliography

Bengio, Samy, and Georg Heigold. 2014. "Word embeddings for speech recognition". In *Fifteenth Annual Conference of the International Speech Communication Association.*

Benzeghiba, Mohamed, Renato De Mori, Olivier Deroo, Stephane Dupont, Teodora Erbes, Denis Jouvet, Luciano Fissore, Pietro Laface, Alfred Mertins, Christophe Ris, et al. 2007. "Automatic speech recognition and speech variability: A review". *Speech communication* 49 (10-11): 763–786. doi:10.1016/j.specom.2007.02.006.

Enarvi, Seppo. 2018. *Modeling Conversational Finnish for Automatic Speech Recognition; Suomen puhekielen mallintaminen automaattista puheentunnistusta varten* [**inlang**en]. 117 + app. 73. Aalto University publication series DOCTORAL DISSERTATIONS; 52/2018. Aalto University; Aalto-yliopisto. ISBN: 978-952-60-7908-0 (electronic); 978-952-60-7907-3 (printed). http://urn.fi/URN:ISBN:978-952-60-7908-0.

Fischer, Hans E, Peter Labudde, Knut Neumann, and Jouni Viiri. 2014. *Quality of instruction in physics: Comparing Finland, Switzerland and Germany.* Waxmann Verlag.

Helaakoski, Jussi, and Jouni Viiri. 2014. "6. Content and Content Structure of Physics Lessons and Students' Learning Gains: Comparing Finland, Germany and Switzerland". *Quality of Instruction in Physics: Comparing Finland, Switzerland and Germany:* 93.

Juang, Biing-Hwang, and Lawrence R Rabiner. 2005. "Automatic speech recognition–a brief history of the technology development". *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara* 1:67.

Kanerva, Jenna, Filip Ginter, Niko Miekka, Akseli Leino, and Tapio Salakoski. 2018. "Turku neural parser pipeline: An end-to-end system for the CoNLL 2018 shared task". *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies:* 133–142.

Silfverberg, Miikka, Teemu Ruokolainen, Krister Lindén, and Mikko Kurimo. 2016. "FinnPos: an open-source morphological tagging and lemmatization toolkit for Finnish". *Language Resources and Evaluation* 50 (4): 863–878. doi:10.1007/s10579-015-9326-3.

Spille, Constantin, Birger Kollmeier, and Bernd T Meyer. 2018. “Comparing human and automatic speech recognition in simple and complex acoustic scenes”. *Computer Speech & Language* 52:123–140. doi:10.1016/j.csl.2018.04.003.

Wieman, Carl, and Katherine Perkins. 2005. “Transforming physics education”. *Physics today* 58 (11): 36. doi:10.1063/1.2155756.