Aleksander Lempinen

Artificial intelligence in physics teacher education

Master's Thesis in Information Technology

May 8, 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: Artificial intelligence in physics teacher education

Työn nimi: Tekoäly fysiikan opettajakoulutuksessa

Project: Master's Thesis

Study line: Educational Technology

Page count: 10+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 Artificial intelligence

TODO: AI overview, AI history, ASR and NLP concepts

2.1 Automatic speech recognition

TODO: ASR overview (Spille, Kollmeier, and Meyer 2018; Enarvi 2018)

2.1.1 Speech to Text

TODO: Different approaches

2.1.2 Finnish speech data

TODO: Conversational vs official Finnish

2.1.3 Comparisons

TODO: Defining "good", comparing literature results, running own tests

2.2 Natural language processing

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

2.2.1 Word embedding

TODO: word2vec

2.2.2 Lemmatization and stemming

TODO: Snowball vs TurkuNLP lemmatization

2.2.3 Finnish NLP data

TODO: Data from different sources, treebanks for parsing vs. word lists/text

2.2.4 Comparisons

TODO: Evaluating the stemming and lemmatization (error rate), evaluating similarities (edit distance, word2vec), literature and own tests

3 Physics instruction analysis

TODO: Physics education (Wieman and Perkins 2005; Fischer et al. 2014)

3.1 Pedagogical link making

TODO: (Scott, Mortimer, and Ametller 2011)

3.2 Conceptual network analysis

TODO: (**vargas**; McLinden 2013; Fischer et al. 2014)

3.3 Methods and materials

3.3.1 Data

TODO: (Fischer et al. 2014)

3.3.2 Speech-to-text

TODO: (Enarvi 2018)

3.3.3 Lemmatization and stemming

TODO: TurkuNLP (Kanerva et al. 2018) TODO: libvoikko https://github.com/voikko/corevoikko/tree/ma
TODO: FinnPOS (Silfverberg et al. 2016)

3.3.4 Word frequency

3.3.5 Network analysis

TODO:

4 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

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.

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.

McLinden, Daniel. 2013. "Concept maps as network data: analysis of a concept map using the methods of social network analysis". *Evaluation and program planning* 36 (1): 40–48. doi:`10.1016/j.evalprogplan.2012.05.001`.

Scott, Phil, Eduardo Mortimer, and Jaume Ametller. 2011. "Pedagogical link-making: a fundamental aspect of teaching and learning scientific conceptual knowledge". *Studies in Science Education* 47 (1): 3–36. doi:`10.1080/03057267.2011.549619`.

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`.