Aleksander Lempinen

Machine learning based automatic analysis of physics instruction quality

Master's Thesis in Information Technology

March 3, 2020

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: Machine learning based automatic analysis of physics instruction quality

Työn nimi: Koneoppimispohjainen automaattinen fysiikan opetuksen laadun analyysi

Project: Master’s Thesis

Study line: Educational Technology

Page count: 18+0

Abstract: TODO: Abstract

Keywords: TODO: Keywords

Suomenkielinen tiivistelmä: TODO: Tiivistelmä

Avainsanat: TODO: Avainsanat

Glossary

TODO	TODO: Glossary
ASR	Automatic Speech Recognition
HMM	Hidden Markov Model
LSTM	Long Short-Term Memory
ML	Machine Learning
NLP	Natural Language Processing
RNN	Recurrent Neural Network

List of Figures

Figure 1. Teacher talk analysis has a common manual pipeline. Classroom video is transcribed, meaningfully coded and then visualized and interpreted. 4

List of Tables

Table 1. The sometimes inaccurate transcript consists of start and end times of the split and the utterance spoken by the teacher during that split 7

Contents

1 INTRODUCTION 1

2 PHYSICS INSTRUCTION QUALITY 3

3 SPEECH AS DATA 5
3.1 Audio data 5
3.2 Text data 6
3.3 Finnish language 6

4 METHODS 7
4.1 Dataset 7
4.2 Natural Language Processing 7
4.3 Network Analysis 7
4.4 Time Series analysis 7
4.5 Supervised machine learning 7

5 RESULTS AND CONCLUSIONS 8

6 RELATED WORK 9

7 DISCUSSION 10

BIBLIOGRAPHY 11

1 Introduction

Automating the laborious and subjective process of analysing classroom speech data has potential for improving physics instruction quality and providing new tools for physics teacher education. Technological advancements in automatic speech recognition (ASR), natural language processing (NLP) and machine learning (ML) allow for more data driven approaches. The aim of this research is to develop techniques for automatic analysis of physics instruction quality.

An important part of physics education is teaching students how to think like an expert and have deeper understanding of the physics concepts with monitoring and guidance of the instructor (Wieman and Perkins 2007). Good physics instruction dependent on the talk between the teacher and the students (Scott and Ametller 2007). Unfortunately teacher talk is not given the attention it deserves during teacher education (Crespo 2002; Lehesvuori 2013), mostly because of lack of good methods to analyse the talk, which is often done with manual labelling or transcribing into text from video and audio data. Analysing teacher talk is discussed in more detail in chapter 2. This presents an opportunity for analysing teacher talk automatically using ASR and NLP methods.

Analysing speech data is particularly difficult due to the variation introduced by the differences between individual speakers, genders, emotional states, accents, dialects or due to casual speech slurring (Benzeghiba et al. 2007). Analysis of the words and the meaning of the speech data will be highly dependent on the available corpus data and tools such as parsers, which were not available for Finnish until recently (Haverinen 2014; Enarvi 2018). Automatic speech recognition and natural language processing is discussed in chapter 3.

In this research we answer the research question "How can traditional NLP techniques and machine learning be used for analysing physics instruction quality?" using the knowledge discovery in databases (KDD) methodology with the following contributions:

- Comparing the quality of lemmatization using the morphological parser Voikko with stemming using the Snowball stemmer when applied to Finnish lesson transcripts in **??**.

- Visualizing the lessons in a meaningful way using network analysis and comparing them using network measures in **??**.
- Engineering features to encode representations of the physics lesson transcripts for use in machine learning models in **??**.
- Applying supervised and unsupervised machine learning techniques to gain insights of the instruction quality in **??**.
- Validation of models and comparison of performance in **??**.

2 Physics instruction quality

The quality of physics instruction has been at the centre of discussion, where traditional teaching methods have been criticized as inefficient at creating experts capable of thinking like physicists (Wieman and Perkins 2007). The interaction between the teacher and the students is called *teacher talk* and is a big part of all physics lessons, which is often taken for granted (Scott and Ametller 2007). Teacher talk is often not sufficiently addressed during teacher education and could be explained by the scarcity of available methods (Lehesvuori 2013; Viiri and Saari 2006; Crespo 2002). This chapter will review the types of methods typically used to analyse teacher talk.

Qualitative observation, where a mentor teacher makes observations during the lesson and gives feedback to the student teacher after the lesson, is an easy and a natural way to analyse teacher talk. According to Viiri and Saari (2006) student teachers have issues with remembering what happened during the lesson and both self-reflection and teacher tutor feedback are based on memory and are unstructured. Therefore they developed a method to analyse teacher talk by visualizing talk types such as "teacher presentation", "authoritative discussion", "dialogic discussion", "peer discussion" and "other" of the lesson. This was achieved by first videotaping the lesson, manually transcribing it and then manually coding each time window into one of the teacher talk types. Viiri and Saari (2006) point out that "Discourse analysis necessarily proceeds on the basis of the investigator's interpretations of what was said." Even this newly developed method is qualitative and subjective in nature by necessity.

This kind of qualitative and subjective analysis of teacher talk from classroom videos, usually by first transcribing or coding the talk based on some kind of a theoretical framework is not unique. For example Scott and Ametller (2007) and Scott et al. (2011) rely on small case studies and manual analysis of teacher talk from classroom video data either directly or from transcripts. The details of what they are looking for in the teacher talk is different in each case and dependent on the chosen theoretical framework, but the overall process is the same.

Qualitative analysis of teacher talk is not necessarily a weakness. Lehesvuori (2013) used a mix of qualitative and quantitative methods in his PhD dissertation and notes that qualitative

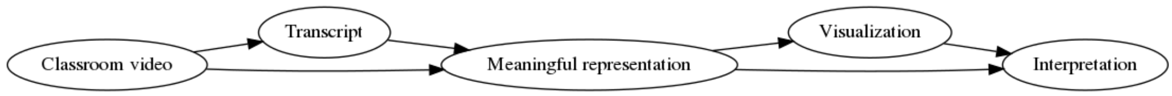

Figure 1. Teacher talk analysis has a common manual pipeline. Classroom video is transcribed, meaningfully coded and then visualized and interpreted.

methods of analysing teacher talk have allowed for more flexibility in contrast with quantitative methods, which were limited by the scarcity of dialogic interactions during the lessons. He identifies a weakness with quantitative methods of analysing teacher talk, which typically do not take into account the temporal aspect of teacher talk continuously changing during the lesson. Lehesvuori (2013) also developed a method to visualize the teacher talk from classroom video of the lesson similar to Viiri and Saari (2006) under a different theoretical framework.

A more quantitative approach was used by Helaakoski and Viiri (Helaakoski and Viiri 2014) to analyse the content structure of teacher talk. They used a relatively large dataset of classroom video consisting of 45 German, 28 Swiss and 25 Finnish lessons about "Relation between electrical energy and power." The videos were manually transcribed and from the videos and transcripts links between concept categories were identified and represented as a connectivity matrix, which could be visualized as a network of concepts. From the same connectivity other metrics and measures were computed. Helaakoski and Viiri (2014) found that "More specifically, the frequencies of physics concepts and connections between them correlated significantly with learning gains." This result is in like with previous research, which stresses the importance of paying attention to teacher talk. (**viiriTeacherTalkPatterns2006** ; Scott and Ametller 2007; Scott, Mortimer, and Ametller 2011)

A common theme in analysing teacher talk is that the pipeline illustrated in figure 1 is manually done. This is obviously a laborious manual task that is prone to errors and is impractical in day-to-day teacher education. While analysing dialogic interactions in teacher talk require interpretation of what the teacher meant (Viiri and Saari 2006), analysing content structure in teacher talk was less subjective and was not as heavily based on interpretation of what the teacher meant (Helaakoski and Viiri 2014). This makes the content structure of teacher talk a prime candidate for automatic analysis of physics instruction quality using NLP methods.

3 Speech as data

3.1 Audio data

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 et al. 2017). 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 et al. 2017).

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 et al. 2017). Speech is highly variable even with a single speaker due to noise, but different pronunciations 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 pronunciations 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 et al. 2017).

3.2 Text data

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

3.3 Finnish language

TODO: Conversational vs official Finnish Finnish language is particularly difficult for ASR because words are formed by concatenating smaller (Enarvi et al. 2017)

4 Methods

4.1 Dataset

The data set of of 25 Finnish physics double lessons on the topic of "Relation between electrical energy and power" was collected during the Quality of Instruction in Physics (QuIP) project (Fischer et al. 2014; Helaakoski and Viiri 2014). The aim of the QuIP project was to:

1. “What are the typical patterns of physics instruction?”
2. “Under what conditions are these patterns successful with respect to student learning, interest and motivation?”

The data consists of transcripts of teacher speech generated by AaltoASR (TODO ask Kurimo/Mansikkaniemi/Virtanen about this), student test results from before and after the lesson and a list of keywords obtained from a physics book glossary. The lessons were conducted by the same teacher wearing a microphone in two roughly 45 minute parts. The generated transcripts are in 5 second splits as shown in table 4.1.

4.2 Natural Language Processing

4.3 Network Analysis

4.4 Time Series analysis

4.5 Supervised machine learning

Start	End	Text
10.0	15.0	hehkulamppua kakstoista tunteesta kirjottaas taas että miten tuo laske
15.0	20.0	virtapiirejä virtapiirejä rakettien kans vaikka kuinka paljon tähän

Table 1. The sometimes inaccurate transcript consists of start and end times of the split and the utterance spoken by the teacher during that split

5 Results and conclusions

TODO:

6 Related work

Automatic speech recognition and NLP techniques have been applied to analyse teacher talk in Spanish and Finnish in a pilot study to determine the feasibility of the approach and maturity of modern ASR systems (Caballero et al. 2017). Two teachers were equipped with microphones and concept networks were created and visualized by counting which physics concept keywords occured together in the teacher's talk. We expanded upon this idea by using a larger dataset with more teachers, using alternative NLP methods such as lemmatization in addition to stemming, adding more representations of the data and including the temporal aspect of the lesson in the analysis and visualizations.

7 Discussion

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 *Proceedings of the 15th Conference of the International Speech Communication Association, Interspeech.*

Benzeghiba, M., R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, et al. 2007. "Automatic Speech Recognition and Speech Variability: A Review". *Speech Communication,* Intrinsic Speech Variations, 49, number 10 (): 763–786. ISSN: 0167-6393, visited on December 15, 2019. doi:10.1016/j.specom.2007.02.006. http://www.sciencedirect.com/science/article/pii/S0167639307000404.

Caballero, Daniela, Roberto Araya, Hanna Kronholm, Jouni Viiri, André Mansikkaniemi, Sami Lehesvuori, Tuomas Virtanen, and Mikko Kurimo. 2017. "ASR in Classroom Today: Automatic Visualization of Conceptual Network in Science Classrooms". In *Data Driven Approaches in Digital Education,* edited by Élise Lavoué, Hendrik Drachsler, Katrien Verbert, Julien Broisin, and Mar Pérez-Sanagustín, 541–544. Lecture Notes in Computer Science. Cham: Springer International Publishing. ISBN: 978-3-319-66610-5. doi:10.1007/978-3-319-66610-5_58.

Crespo, Sandra. 2002. "Praising and Correcting: Prospective Teachers Investigate Their Teacherly Talk". *Teaching and Teacher Education* 18, number 6 (): 739–758. ISSN: 0742-051X, visited on February 8, 2020. doi:10.1016/S0742-051X(02)00031-8. http://www.sciencedirect.com/science/article/pii/S0742051X02000318.

Enarvi, Seppo. 2018. *Modeling Conversational Finnish for Automatic Speech Recognition.* Aalto University. ISBN: 978-952-60-7908-0, visited on December 16, 2019. https://aaltodoc.aalto.fi:443/handle/123456789/30638.

Enarvi, Seppo, Peter Smit, Sami Virpioja, and Mikko Kurimo. 2017. "Automatic Speech Recognition with Very Large Conversational Finnish and Estonian Vocabularies" (). Visited on December 16, 2019. doi:10.1109/TASLP.2017.2743344. https://arxiv.org/abs/1707.04227v5.

Fischer, Hans E, Peter Labudde, Knut Neumann, and Jouni Viiri. 2014. *Quality of Instruction in Physics: Comparing Finland, Germany and Switzerland.* Münster: Waxmann Verlag. ISBN: 978-3-8309-3055-6.

Haverinen, Katri. 2014. "Natural Language Processing Resources for Finnish. Corpus Development in the General and Clinical Domains" (). Visited on December 16, 2019. `https://www.utupub.fi/handle/10024/98608`.

Helaakoski, Jussi, and Jouni Viiri. 2014. "Content and Content Structure of Physics Lessons and Students' Learning Gains: Comparing Finland, Germany and Switzerland". In *Quality of Instruction in Physics: Comparing Finland, Germany and Switzerland,* 93–110. Münster: Waxmann Verlag. ISBN: 978-3-8309-3055-6.

Juang, B H, 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* 1:24.

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". In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies,* 133–142. Brussels, Belgium: Association for Computational Linguistics. Visited on December 16, 2019. doi:`10.18653/v1/K18-2013`. `https://www.aclweb.org/anthology/K18-2013`.

Lehesvuori, Sami. 2013. "Towards Dialogic Teaching in Science : Challenging Classroom Realities through Teacher Education". *Jyväskylä studies in education, psychology and social research,* number 465. Visited on January 26, 2020. `https://jyx.jyu.fi/handle/123456789/41268`.

Lehesvuori, Sami, Jouni Viiri, Helena Rasku-Puttonen, Josephine Moate, and Jussi Helaakoski. 2013. "Visualizing Communication Structures in Science Classrooms: Tracing Cumulativity in Teacher-Led Whole Class Discussions". *Journal of Research in Science Teaching* 50 (8): 912–939. ISSN: 1098-2736, visited on February 6, 2020. doi:`10.1002/tea.21100`. `https://onlinelibrary.wiley.com/doi/abs/10.1002/tea.21100`.

Scott, Phil, and Jaume Ametller. 2007. "Teaching Science in a Meaningful Way: Striking a Balance between 'opening up' and 'Closing down' Classroom Talk".

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, number 1 (): 3–36. ISSN: 0305-7267, visited on December 16, 2019. doi:10.1080/03057267.2011.549619. https://doi.org/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, number 4 (): 863–878. ISSN: 1574-0218, visited on December 16, 2019. doi:10.1007/s10579-015-9326-3. https://doi.org/10.1007/s10579-015-9326-3.

Viiri, Jouni, and Heikki Saari. 2006. "Teacher Talk Patterns in Science Lessons: Use in Teacher Education". *Journal of Science Teacher Education* 17, number 4 (): 347–365. ISSN: 1573-1847, visited on February 6, 2020. doi:10.1007/s10972-006-9028-1. https://doi.org/10.1007/s10972-006-9028-1.

Wieman, Carl, and Katherine Perkins. 2007. "Transforming Physics Education". *Physics Today* 58, number 11 (): 36. ISSN: 0031-9228, visited on December 16, 2019. doi:10.1063/1.2155756. https://physicstoday.scitation.org/doi/abs/10.1063/1.2155756.