

Bachelor thesis - Better accuracy of automatic lecture transcriptions by using context information from slide contents

Jonathan Werner

1 Introduction

Scannability is crucial for academic research: you have to be able to quickly evaluate the usefulness of a given resource by skimming the content and looking for the parts that are specifically relevant to the task at hand.

The medium in which those resources are available is very centered on textual representation. Spoken content, hereinafter called **speech media** (audio- or audiovisual media that mainly consists of spoken language) **doesn't make it possible to scan its contents**. You are "stabbing in the dark" when looking for something specific in a medium like this and have to consume it like a linear narrative.

This means that although lectures and conference talks are a central element to science they are much more challenging and tedious to use for research work.

Being able to a) efficiently **search** and b) look at the **temporal distribution of important keywords** in a visually dense way would elevate the usefulness of speech media in the scientific context immensely.

One approach to accomplish those goals is utilizing Automatic Speech Recognition (ASR) to transcribe speech to text and also get timing information for the recognized words. This makes it possible to derive information about the density of given words at a given point of time in the talk, which in turn allows to compute **word occurrence density maxima**. This opens up possibilities for **compact visual representation** of the interesting keywords, thus allowing the user to **scan**.

The main challenge when using ASR for this task is the recognition accuracy of technical terms. Most of them are not included in the language models that are available as those are broad and generic so as to optimize for accuracy over a wide topic spectrum. But when they are not included into the language model they have no chance to be correctly recognized at all.

So the usefulness of applying ASR with a generic language model to the problem is very small, as the interection of interesting keywords with those technical terms that can not be recognized is very big.

The central goal of this thesis is to explore an approach to overcome this problem. This approach consists of using words from lecture slides or other notes to **generate a lecture-specific language model**. This is then **merged** with a generic language model and being compared to the ‘baseline’ accuracy of the generic model.

1.1 Summarize research questions

2 Background

3 Methodology

4 Process

5 Discussion and Findings

6 Improvements

7 Conclusions

Bibliography