Report: Dialog Act Classification
using Word Embeddings & Acoustic Features

Jens Beck jens.beckl@ims
Fabian Fey fabian.fey@ims
Richard Kollotzek richard.kollotzek@ims

Abstract

1 Introduction

The general task is to classify lexical and auditory speech into one of four predefined *dialog act classes*. A *dialog act*, in this context, represents informal information of how a dialog system should respond to a users input. The four provided classes are *statement, opinion, question* and *backchannel*. To solve this task we developed *convolutional neural networks* (CNN) that use lexical and acoustic features. For the development and training of the systems a subset of the *Switchboard Dialog Act Corpus* was used. In next chapters we discuss the development of the systems and subsequently to that the research question **INSERT HERE**.

2 Data & Data Preperation

In this section we discuss the *Switchboard Dialog Act Corpus* and the extraction of the lexical and acoustic features.

2.1 The Switchboard Dialog Act Corpus

The *Switchboard Dialog Act Corpus*, from now on abbreviated as *SwDA*, consists of recordings with corresponding transcripts and *dialog act classes*.

2.2 Data Preprocessing

3 Baseline Systems

4 Results

5 Research Question: None

6 Conclusion

References

Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: The munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pages 1459–1462, New York, NY, USA, 2010. ACM.

Yoon Kim. Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882, 2014.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546, 2013.