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

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**.

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

The *Switchboard Dialog Act Corpus* (0), from now on abbreviated as *SwDA*, consists of recordings with corresponding transcripts. Each of these recordings is assigned to one of 42 *dialog act classes*. In this project we reduced the amount of classes down to four which are *statement, opinion, question* and *backchannel*. These classes are supersets of the 42 *dialog act classes* defined in the *SwDA*.

References

J. J. Godfrey, E. C. Holliman, and J. McDaniel. Switchboard: Telephone speech corpus for research and development. In *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing - Volume 1*, ICASSP'92, pages 517–520, Washington, DC, USA, 1992. IEEE Computer Society.