

Dialog Act Classification

using Word Embeddings & Acoustic Features

Jens Beck, Fabian Fey, Richard Kollotzek
Institute for Natural Language Processing, University of Stuttgart



Task Introduction

-
-
- Short examples:
 - What is your name? - Question
 - It's raining. - Statement
- Dialog Act Classification
- What are Dialog Acts
- Short examples

Data

Switchboard

- Subset of the Switchboard Telephone Speech Corpus
- The Switchboard Corpus has lexical and acoustic data
- The lexical dataset are divided in training-, development- and test-sets

Dataset\Channel	opinion	question	backchannel	statement	Sum
training	4984	2150	6792	14459	28385
development	1068	460	1455	3098	6081
test	1070	463	1458	3099	6090

- The acoustic dataset includes a recording of every utterance

MFCC features

- Extraction of the MFCC features for every sentence with OpenSmile
- Every 25ms the MFCC features where extracted, which resulted in 13 features for each measurement point

word2vec

- For the word embedding layer we used the Google word2vector model
 - It contains 3 million words and phrases with a 300-dimensional vector each

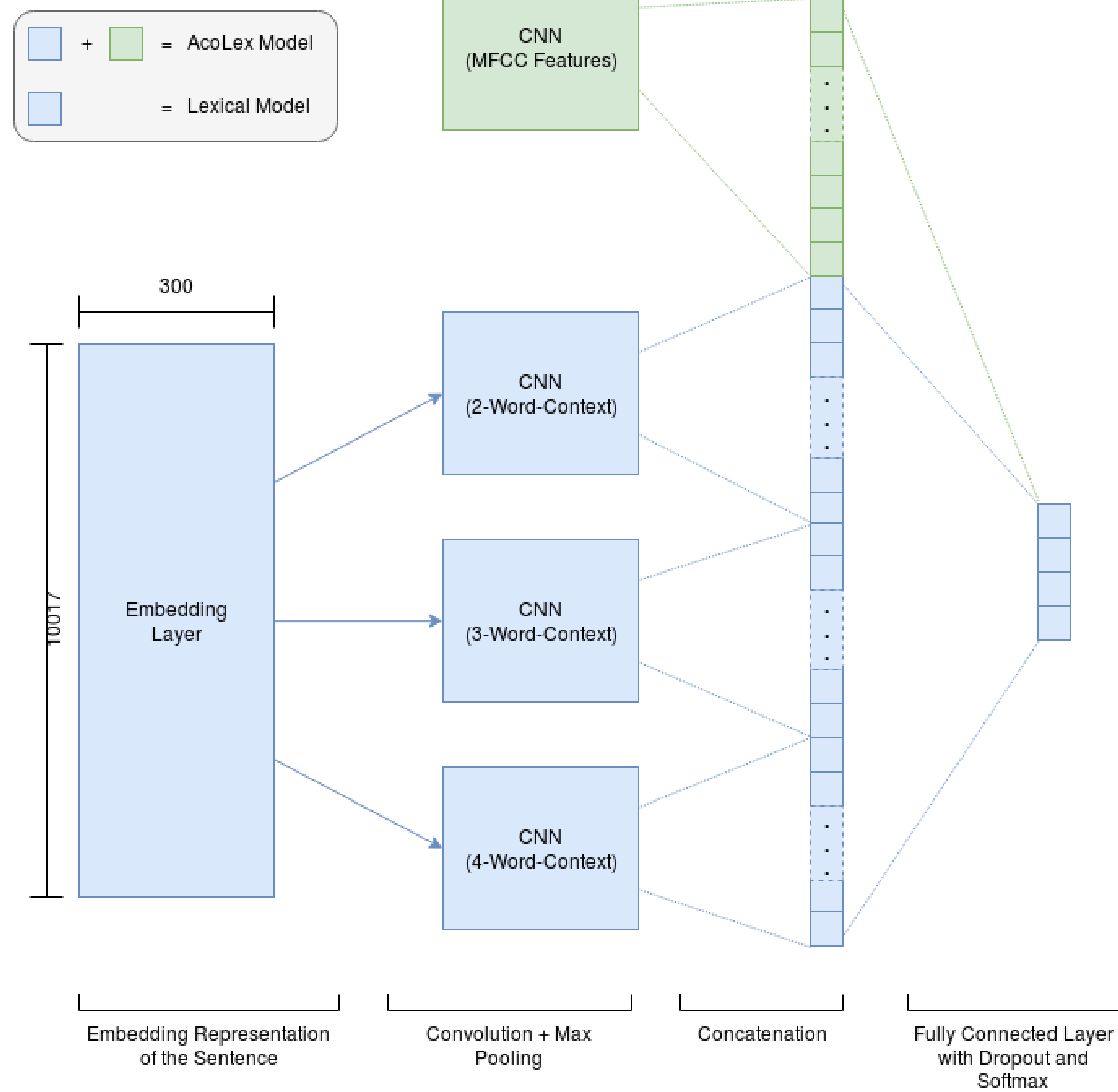
Data Preprocessing

- All words from the three datasets are indexed in a word list for use in the embedding matrix
 - Each word in word list, from the training set, was given the corresponding vector from the word2vec model
 - A random vector is assigned if the word is not in the word2vec model or from the test and development set
 - Each sentence is converted to a sequence with the corresponding indexes from the word list
- The MFCC feature matrix is reduced to the first 1000 and the last 1000 measurement points, which results in a 13 by 2000 matrix

1. asdafaoi asdihj osadh sad

System Architecture

- Architecture visualization



Intermediate Results

- Table with different configs and mean accuracies

Potential Future Work

- Further optimizations
- Maybe varying MFCC feature size
- Including word2vec features for the test and development set
- Introducing additional layer between CNN output and softmax