Research Week 2020

Automatic topic segmentation of ADOS transcripts

Joel Adams, Steven Bedrick

OHSU

Keywords

autism, machine learning, natural language processing

Abstract

Introduction

The Autism Diagnostic Observation Schedule (ADOS) is a widely used instrument for the diagnosis of Autism Spectrum Disorder (ASD). It includes tasks designed to elicit conversational language between the subject and an examiner.

Atypical language use and repetitive speech are features of the disorder as characterized by the DSM-5. Quantifying what is atypical about the language of subjects with ASD is challenging.

To answer questions like "Are subjects more or less verbose when discussing loneliness than their typically developing peers?" researchers have to manually label large sections of the ADOS. This is exceptionally time consuming.

We applied statistical sequence tagging together with state-of-the-art Natural Language Processing (NLP) tools to leverage the structure of the ADOS and automatically label each sentence with its framing question.

Methods

Our data consists of 115 ADOS module-3 transcripts (64 TD, 51 ASD) from children in an autism study. We manually annotated a randomly selected subset of 40.

ADOS conversations are guided by a set of scripted questions the examiner asks the subject. A subject might be asked "What makes you feel happy?" or "What are the things you're afraid of?" We marked each utterance with a tag representing which of these questions it was related to.

We created features from vector representation of the words in each utterance, and the similarity of the previous examiner utterance to each of the scripted questions (using the BERT vector semantic model). We trained a conditional random field (CRF) model on the tags and features, which can then predict tags on an unlabeled transcript.

Results

Tags generated using our model appear reasonable (f1 = .87, sd=.06). Preliminary tests suggest they generalize well to unlabeled transcripts. This work will allow us to perform new analyses via NLP and other computational methods.