DETECTING LANGUAGE IMPAIRMENTS IN AUTISM: USING SEMI-STRUCTURED CONVERSATIONS WITH VECTOR SEMANTICS

ADAM GOODKIND  MICHELLE LEE  GARY E. MARTIN
MOLLY LOSH     KLINTON BICKNELL

NORTHWESTERN UNIVERSITY & ST. JOHN'S UNIVERSITY

SOCIETY FOR COMPUTATION IN LINGUISTICS
• Earliest diagnostics autism has been associated with atypical language and communication (Kramer, 1943)

• Descriptions of speech:
  • “peculiar and out of place in ordinary conversation”
  • “stereotyped”
  • “inappropriate”
  • “a lack of ease in the use of words”

• Rely on “borrowed” utterances from others

• Non-developmental syntax errors
  • E.g. echoing pronouns
PRIOR WORK ON DETECTING ASD

- Many prior studies have evaluated the semantics of autism
  - Losh & Capps, 2003
  - Prud'hommeaux et al., 2011
  - Rouhizadeh, Prud’hommeaux, van Santen, & Sproat, 2014
- Had children retell a narrative: story, movie, cartoon, etc.
  - Compared to an objective gold standard, i.e. the original story/movie
GOALS

• Differentiate language of children with autism from language of typically developing children using language from semi-structured activities

• How?
  • Quantify semantic language from these activities using word vectors
  • Measure semantic similarity of each child’s language to a gold standard of typical development
  • Assess new confound: Explore how semantics is affected by transcript length
OUR SUBJECTS

• Control Group
  • Typically developing (TD)

• Clinical Group
  • Autism Spectrum Disorder (ASD)
  • Fragile X Syndrome (FXS)
  • Fragile X Syndrome Comorbid with ASD (FXS-A/FXS-ASD)
THE DATASET

- 109 total subjects
  - All male
- Controlled for mental age

<table>
<thead>
<tr>
<th></th>
<th>Chronological Age</th>
<th>Mental Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Clinical Group</td>
<td>~9</td>
<td>5</td>
</tr>
</tbody>
</table>
SESSION ACTIVITIES

• Session activities outlined in ADOS
  • A benchmark diagnostic test for ASD (Lord, et al., 2000)
  • Conversations
    • Building a contraption
    • Playing with dolls/action figures
    • Less structured
METHODOLOGY

TRANSCRIPTS

• All sessions transcribed, with pause markings and Xed out unintelligible words

• Two control transcripts (TD) were designated “Gold Standards”
  • Selected by 2 researchers, based on clinical-behavioral ratings
    • Minimal pragmatic language impairments
    • High scores on core conversational metrics
      • E.g. contingency, reciprocity, initiation
  • Gold standard transcripts not included in the TD group
METHODOLOGY
TRANSCRIPT PROCESSING

• Stop words were *not* removed
  • Personal pronouns, definitives, etc.
  • Improved results, but did qualitatively change them
• Child utterances converted to 400-dimension word vectors
WORD VECTORS

• Vector of weights representing dimensions of a word
• Each word vector representation is derived from co-occurrence count with other words

<table>
<thead>
<tr>
<th>Word</th>
<th>Vector Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>boat</td>
<td>[5, 1]</td>
</tr>
<tr>
<td>banana</td>
<td>[2, 2]</td>
</tr>
<tr>
<td>apple</td>
<td>[2, 3]</td>
</tr>
</tbody>
</table>

• Comparison of vectors (cosine distance)
WORD VECTORS → TRANSCRIPT VECTORS

• Used Google’s word2vec pre-trained word vectors
  • Trained on newswire text

• Transcript vector creation process
  • Each word (in vocabulary) converted to a vector
  • Add all word vectors together
  • Normalize final vector

• Used vector of entire transcript for comparison of semantic similarity
METHODS FOR COMPARISON

• Establish a mean vector from Gold Standards against which to measure similarity
• Compared entire transcript vector of each child to the mean vector
  • Future studies will also compare questions to answers within a transcript
INITIAL SEMANTIC SIMILARITY RESULTS

- TD has highest semantic similarity
- ASD and FXS-A had lower semantic similarity than TD
- Possible explanations:
  - Unconventional choice of topics
  - Should be apparent in a semantic model
  - Unconventional word choices
  - Synonyms should not make a difference
MEASURING VARIABILITY

• Reduce every transcript vector to two most informative dimensions, using PCA
• Measure size/density of diagnostic group clusters
• TD subjects are closely clustered together
• ASD subjects much more diffuse
• Children with ASD also differ more from each other
SUMMARY OF INITIAL RESULTS

• Significant differences in semantic similarity to gold standard between control and clinical groups
• Semantic variability within clinical groups are also more pronounced
POSSIBLE CONFOUND
TRANSCRIPT LENGTHS

• Transcript length is not uniform across diagnostic groups
• Subjects with longer transcripts have higher semantic similarity scores
TRANSCRIPT LENGTHS

• Could be that more advanced language users talk for longer periods of time
  • Not problematic for our results
• Could be that the semantic similarity measure is biased higher for longer transcripts (e.g., averaging across more words just yields more robust estimates of the mean)
  • In this case, differences between groups in similarity to the gold standard might be driven solely by length differences
SEMANTIC SIMILARITY RESULTS
CONTROL FOR TRANSCRIPT LENGTH

- Randomly sample fixed number of words from all transcripts
- Sample size set at 300 words
  - 98% of transcripts longer than 300 words
- Negates the effect of length
- We still find that TD is significantly more similar to a gold standard than ASD and FXS-ASD
• Subset relationship still holds when controlling for transcript length
• Distance between points (cluster density) remains the same as using all words
CONCLUSIONS

• Developmental disorders such as ASD have linguistic signals that can be detected in less controlled/structured environments
• Vector semantics can be used to detect linguistic differences associated with ASD
• Computational linguistics has the potential to improve clinical diagnostics
THANK YOU

Funding sources:

Adam Goodkind
a.goodkind@u.northwestern.edu