Screening for Alzheimer's with psycholinguistics

Marten van Schijndel (The Ohio State University) & Kathleen C. Fraser (University of Toronto)
vanschm@ling.osu.edu

Introduction Accurate and non-invasive screening for Alzheimer's disease (AD) is critical to allow patients time to plan for the future and access early treatment. The present work studies the effectiveness of well-known psycholinguistic measures at detecting likely cases of AD from narrative speech. Since AD is correlated with memory impairment, this study tests a measure of linguistic memory load (embedding depth) and a measure of changes to working memory load (embedding difference [5]), and since memory impairment can have linguistic consequences this study tests an information-theoretic measure of processing complexity (surprisal).

Methods This study uses the publicly available DementiaBank corpus [1], which contains narrative speech elicited through a picture description task. Subjects with a diagnosis of "Possible" or "Probable" AD were assigned to a single AD group (n=167); healthy, elderly subjects were assigned to a control group (n=98). For this study, half the subjects of each group were included in a development partition for data exploration while the remainder were used for significance testing.

Detecting Alzheimer's Logistic mixed regression was used to detect AD at the word level with a random intercept for each word, by-word random slopes for sentence position, 5-grams (from Gigaword 4.0 [3]), surprisal, embedding depth, and embedding difference, and the following baseline fixed effects: sentence position, word length, unigram frequency (obtained from SUBTL [2]), and all 2-way interactions. Any words absent from SUBTL were removed from the analysis. Adding a fixed effect of 5-grams to the baseline produced significantly better classification accuracy ($p < 0.001$) as did the subsequent addition of embedding depth ($p < 0.001$).[1] Surprisal ($p = 0.11$) and embedding difference ($p = 0.23$) failed to improve the model when 5-grams were present.

Coefficient analysis suggests that subjects with AD use shorter sentences ($p < 0.001$), more deeply embedded phrases ($p < 0.001$), and more common words ($p < 0.001$) which tend to be in unusual lexical contexts (lower 5-gram probability, $p < 0.001$). The shorter sentences and more common words suggest an overall reduction in the linguistic complexity of AD speech. The finding that subjects with AD produce deeper embeddings than controls was unexpected since it suggests subjects with AD produce structures with higher memory cost, but the effect seems driven by frequent parentheticals and asides caused by distraction during AD narratives. The low 5-gram probability in the AD group seems to be driven by a tendency for more telegraphic speech.

While these findings indicate that the impaired memory of AD subjects affects their surface lexical distributions, the weakness of surprisal as a predictor suggests the underlying syntactic distributions are relatively unaffected. Further, the unhelpfulness of embedding difference suggests that updating working memory is not more costly for AD subjects than controls, which implies that the memory difficulties in AD stem from storage or access difficulty.

Results This work shows that psycholinguistic measures of frequency and memory load are robust predictors of AD. They can be easily applied to any linguistic output generated by those suspected of having AD. In practice, these measures may only be applicable to narratives, but since traditional diagnostic tests for AD such as the Wechsler Memory Scale [4] involve narrative components, the measures from this study can cheaply and easily augment other diagnostic data as it is gathered.

References

[1] Francois Boller and James Becker. *DementiaBank Pitt Dementia Corpus*, 2005.
[2] Marc Brysbaert and Boris New. Moving beyond Kucera and Francis. *Behavior Res. Methods*, 41(4):977–90, 2009.
[3] Robert Parker, et al. *English Gigaword LDC2009T13*, 2009.
[4] David Wechsler. *Wechsler Memory Scale - Third Edition Manual*. The Psychological Corporation, 1997.
[5] Stephen Wu, et al. Complexity metrics in an incremental right-corner parser. In *ACL 2010*, pages 1189–1198, 2010.

[1] Significance for model improvement comes from ablative ANOVAs between each model and the next simpler one.