Propositional idea density and other linguistic features for predicting Alzheimer's disease from transcribed interviews

Kairit Sirts, Mark Johnson

Macquarie University

NeuRA lab meeting 08.06.2016

Outline

1. Alzheimer's disease and language
2. Propositional idea density
3. Experimental results

Outline

Alzheimer's disease

- Alzheimer's disease (AD) is the most common form of dementia.
- AD damages the brain - impaired memory, thinking and behavior
- The biggest risk factor for AD is age
- There is no single test for diagnosing AD - various clinical and neuropsychological tests are conducted to rule out other possible diseases and to make a tentative diagnosis
- The diagnosis can only be confirmed after death via brain examination
- There is no cure for AD

Language disfluency is one of the earliest symptoms of AD

- Changes in lexical content:
  - The lexicon becomes smaller and more general
  - The rate of pronouns increases
  - More present tense verbs
- "Empty speech"
  - Increased usage of lexical fillers (I mean, you know, etc)
  - Increased usage of formulaic language
  - Repetitive content
  - However, the grammatical abilities do not suffer
- Syntactic complexity decreases - shorter sentences, utterances, clauses
- More pauses, non-lexical fillers (uh, huh, ah)

Motivation for studying linguistic features of AD

- Spontaneous speech is easy to obtain
- Simple and quick methods for monitoring the condition or helping to establish the diagnosis

Goal of this work

Extract linguistic features from the transcribed speech of AD patients (and healthy controls) and train predictive computational models to discriminate between AD and healthy subjects.

Outline

Propositional idea density (Kintsch and Keenan, 1973)

- Rate of semantic propositions or ideas in a text or utterance
- Each proposition can be stated as a question having a yes/no answer

The old gray mare has a very large nose.

1. (HAS, MARE, NOSE)
2. (OLD, MARE)
3. (GRAY, MARE)
4. (LARGE, NOSE)
5. (VERY, (LARGE, NOSE))

Types of propositions

Elementary propositions

- Predications - organized around a verb
  - (HAS, MARE, NOSE)
- Modifications - adjectives, adverbs
  - (LARGE, NOSE)

Complex propositions

- Both predications and modifications
  - (VERY, (LARGE, NOSE))
- Connective propositions
  - Join two propositions with a conjunction or causal relationship
  - I got wet because of rain
  - (BECAUSE, (IS, RAINING), (GET, I, WET))

PID and AD

The Nun study (Snowdon et al. 1996)

- Analyse the essays written by nuns 50 years ago before taking their religious vows.
- All nuns agreed to donate their brains after death for research
- Study a sample of 25 (dead) subjects, 10 of whom were diagnosed with AD
- Result: AD was associated with the lower PID in the youth

Precursors study (Engelman et al. 2010)

- Students entering the John's Hopkins School of Medicine in 1948-1968
- Their admission essays were analysed
- Studied a sample of 54 subjects, 18 with AD
- Result: Confirmed the results of the Nun study

PID and AD

- Propositional idea density measures some sort of cognitive reserve that has a protective effect from Alzheimer's disease. (Chand et al., 2012)

PID in computational studies

Jarrold et al., 2010

- Analysed the transcribed interviews recorded with the cognitively normal subjects
- 22 of the total 55 subjects later died with the death cause marked as AD
- Found that the PID of AD group was significantly lower from the control group
- Classification accuracy using PID + a range of other features 73%

Roark et al., 2011

- Analysed the transcribed story retellings of subjects with mild cognitive impairment
- 37 patients with MCI and 37 healthy controls
- Found no significant difference in PID between groups

Computing PID

- Many studies have used manual counting
- CPIDR program (Brown et al., 2008) - PID is computed automatically based on part-of-speech tags
- Generally propositions correspond to verbs, adjectives, adverbs, prepositions and coordinating conjunctions

Outline

AMI - Autobiographical Memory Interviews

- Interviews conducted and transcribed at UNSW
- 20 AD patients, 20 controls
- Each interview consists of four stories where subjects describe events from different periods of their life: teenage, young adult, middle age, last year

	AMI-Interviews		**AMI-Stories**	
	AD	**Control**	**AD**	**Control**
Samples	36	20	143	80
Words per sample	4743	5422	1193	1355

Training

- Train and test with 10-fold cross-validation
  - Data is divided into 10 non-overlapping subsets (folds)
  - Nine folds are used for training, 1 fold is used for testing
  - Repeat with all training-test fold combinations
- Repeat 100 times and report the mean accuracy.
- Regularised logistic regression
  - Regularization: model does not allow large coefficients
- Perform feature selection using Pearson's correlations and then train models using those features only

PID means of AD and Control groups

Data set	**AD**	**Control**	
AMI-Interviews	0.51 (0.02)	0.51 (0.01)	
AMI-Stories	0.51 (0.03)	0.51 (0.02)	
Jarrold et al., 2010	0.50 (0.02)	0.53 (0.02)	*

PID classification accuracy

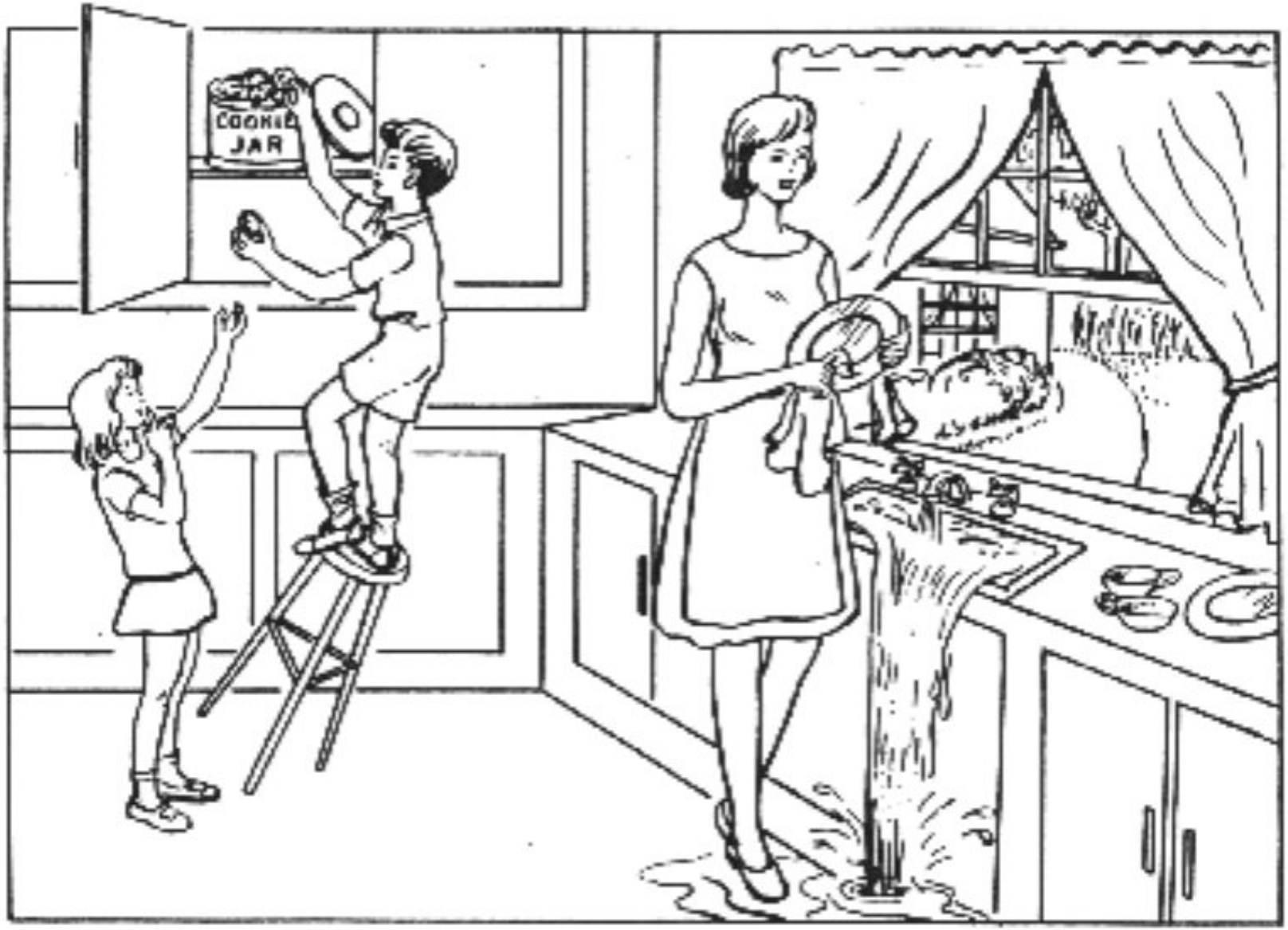
Data set	Majority baseline	PID
AMI-Interviews	64.3	57.5 (2.9)
AMI-Stories	64.1	59.1 (2.4)

Talkbank

- Publicly available data set containing transcribed interviews with patients with dementia and healthy control subjects
- Subjects are asked to describe the Cookie Theft picture
- Some subjects have given several interviews over the years

	AD	**Control**
Subjects	169	98
Samples	257	241
Words per sample	105	114

Cookie Theft picture



PID on Talkbank data

Table: PID means of AD and Control groups

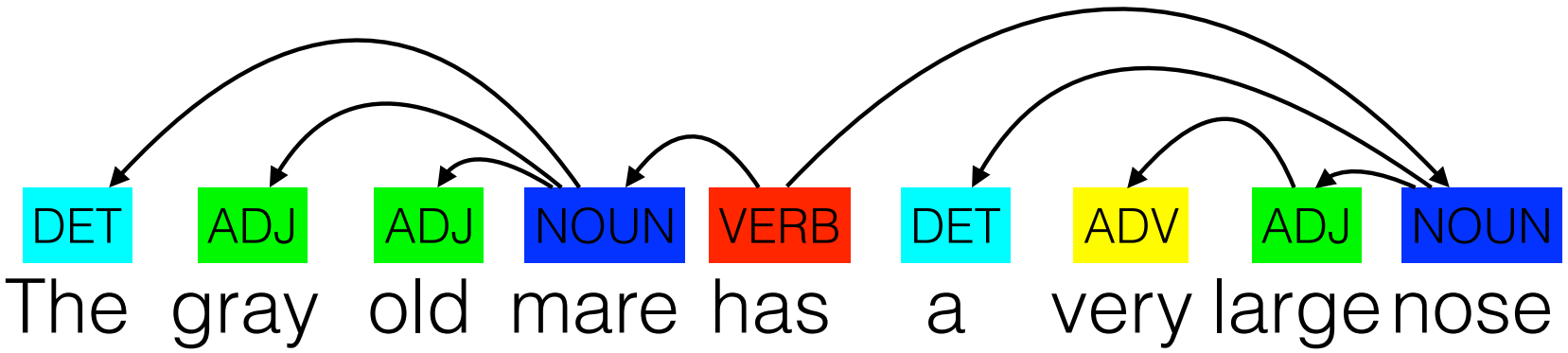
Data set	**AD**	**Control**	
AMI-Interviews	0.51 (0.02)	0.51 (0.01)	
AMI-Stories	0.51 (0.03)	0.51 (0.02)	
Talkbank	**0.53 (0.07)**	**0.50 (0.06)**	***
Jarrold et al., 2010	0.50 (0.02)	0.53 (0.02)	*

Table: Accuracy of logistic regression trained with the PID feature only

Data set	**Majority baseline**	**PID**
AMI-Interviews	64.3	57.5 (2.9)
AMI-Stories	64.1	59.1 (2.4)
Talkbank	**51.5**	**58.8 (0.8)**

Modeling PID with dependency structure

- Dependency parsing is one way of analysing the syntactic structure of sentences.



Similarity of propositional and dependency structures

The old gray mare has a very large nose.	
Dependencies	**Propositions**
det(The, mare)	
amod(old, mare)	(OLD, MARE)
amod(gray, mare)	(GRAY, MARE)
nsubj(mare, has)	(HAS, MARE, NOSE)
det(a, nose)	
advmod(very, large)	(VERY, (LARGE, NOSE))
amod(large, nose)	(LARGE, NOSE)
dobj(nose, has)	(HAS, MARE, NOSE)

POS and dependency features

1. Coarse POS tags: ADJ, VERB, NOUN
2. Fine POS tags: NN, NNP, VBZ, VBG
3. Dependency relation types: amod, nsubj, dobj
4. Coarse POS pairs with dependency type: ADJ_amod_NOUN, NOUN_nsubj_VERB, ADV_admod_ADJ,
5. Fine POS pairs with dependency type: DT_det_NN, JJ_amod_NN, NN_nsubj_VBZ

Classification results with POS and dependency features

Features	**AMI-Interviews**	**AMI-Stories**	**Talkbank**
Coarse POS	56.1 (3.7)	59.3 (2.6)	68.5 (1.0)
Fine POS	56.8 (3.3)	61.7 (2.4)	72.5 (1.0)
Type only	57.3 (2.9)	60.2 (2.5)	71.9 (0.8)
Coarse POS + type	53.2 (3.6)	57.2 (2.3)	71.6 (1.1)
Fine POS + type	55.7 (3.2)	63.0 (2.8)	**75.7 (1.1)**
PID baseline	57.5 (2.9)	59.1 (2.4)	58.8 (0.8)
Majority baseline	**64.3**	**64.1**	51.5

Classification using other features

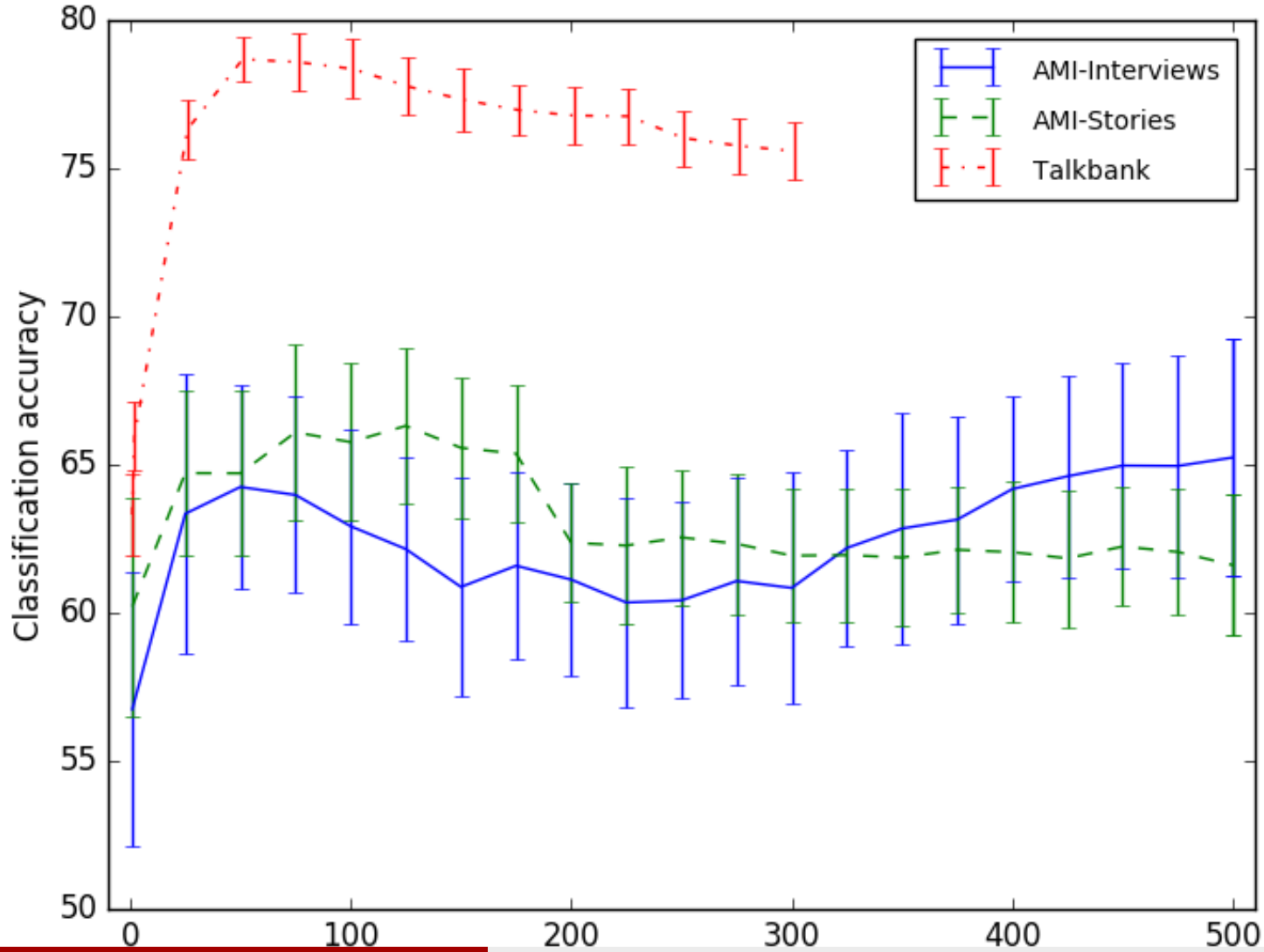
We extracted a range of features proposed by Fraser et al. (2015)

- POS tags and POS tag ratios
- Syntactic features (clauses, T-units, parse tree height, Yngve depth, sentence length)
- Vocabulary richness features (type-token ratio, Honore's statistic, Brunet's index, phrase type counts word length)
- Constituency parse tree rules
- Psycholinguistic features (frequency, familiarity, imageability, age of acquisition)
- Features measuring repetition

Classification results using linguistic features

Features	**AMI-Interviews**	**AMI-Stories**	**Talkbank**
POS	65.8 (3.3)	63.7 (2.4)	72.2 (0.9)
Syntax	**71.1 (2.6)**	**72.4 (1.8)**	69.2 (0.9)
Grammatical	53.1 (3.9)	54.6 (2.6)	75.3 (1.0)
Vocabulary	69.3 (3.7)	66.0 (2.2)	58.8 (1.1)
Psycholinguistic	64.4 (4.0)	68.2 (2.0)	70.4 (0.8)
Repetition	59.3 (2.6)	65.2 (2.0)	54.6 (1.7)
All	53.8 (4.9)	63.1 (2.7)	**76.3 (0.9)**
Majority baseline	64.3	64.1	51.5

Classification with feature selection



Classification with feature selection

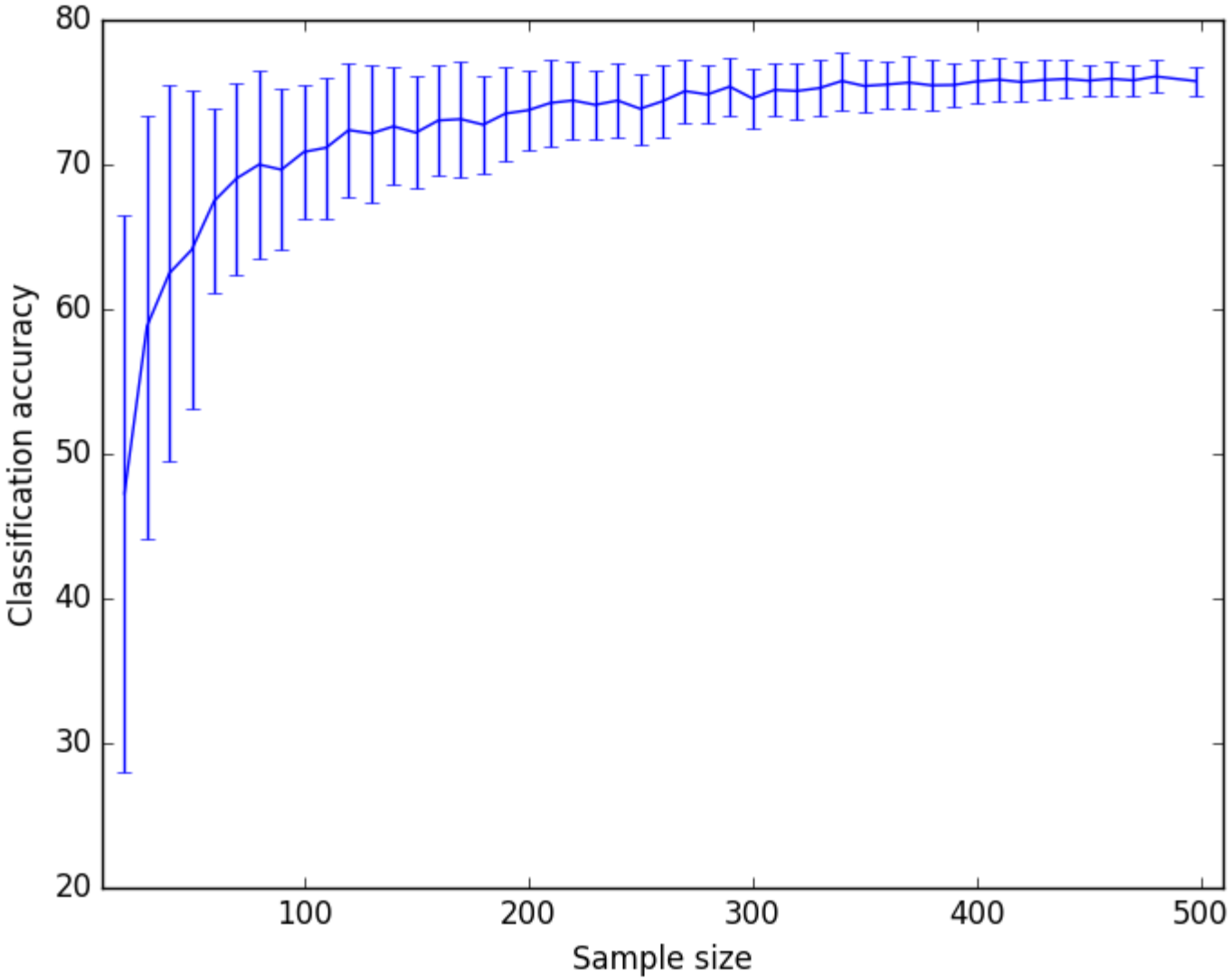
Data set	**# Features**	**Accuracy**
AMI-Interviews	500	65.2 (4.0)
AMI-Stories	125	66.3 (2.6)
Talkbank	50	78.8 (0.8)
(Fraser et al., 2015)	35	81.9
(Fraser et al., 2015)	50	78.7

Subsampling Talkbank data

- N=56, 36 AD, 20 Controls, all features

Data set	**Accuracy**
AMI-Interviews	65.1 (3.7)
AMI-Stories	62.1 (2.1)
Talkbank	66.2 (5.0)

Subsampling the Talkbank data



Summary

- No significant differences in PID on AMI data
- PID significantly different on Talkbank data, but the direction is wrong - AD patients had higher PID
- The classification results on AMI data in most cases worse or barely better than the majority baseline.
  - The sample size is probably too small for training predictive modeling
- We were able to get similar results on Talkbank as Fraser et al., (2015) with a slightly simpler feature set

References

- Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging.
- Chand, V., Baynes, K., Bonnici, L. M., & Farias, S. T. (2012). A rubric for extracting idea density from oral language samples.
- Engelman, M., Agree, E. M., Meoni, L. A., & Klag, M. J. (2010). Propositional density and cognitive function in later life: findings from the Precursors Study.
- Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., & Rochon, E. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts.
- Jarrold, W. L., Peintner, B., Yeh, E., Krasnow, R., Javitz, H. S., & Swan, G. E. (2010). Language analytics for assessing brain health: cognitive impairment, depression and pre-symptomatic Alzheimer's disease.
- Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences.
- Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K., & Kaye, J. (2011). Spoken Language Derived Measures for Detecting Mild Cognitive Impairment.
- Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study.