Automatically Identifying Trouble-Indicating Speech Behaviors in Alzheimer's Disease

Frank Rudzicz[1,2], Leila Chan Currie[2], Andrew Danks[2], Tejas Mehta[2], Shunan Zhao[2]
[1] Toronto Rehabilitation Institute - University Health Network; and
[2] Department of Computer Science, University of Toronto
Toronto, Canada
{frank,leila,danks,tmehta,szhao}@cs.toronto.edu

ABSTRACT

Alzheimer's disease (AD) deteriorates executive, linguistic, and functional capacity and is rapidly becoming more prevalent. In particular, AD leads to an inability to follow simple dialogues. In this paper, we annotate two databases of dyad conversations, that include individuals with AD, with *trouble indicating behaviors* (TIBs). We then extract lexical/syntactic and acoustic features from all utterances and identify those that are most indicative of TIB (which include speech rate and utterance likelihoods in a standard language model) and classify utterances as having TIB or not with up to 79.5% accuracy. This will allow us to build automated dialogue systems and assessment tools that are sensitive to confusion in people with AD.

Categories and Subject Descriptors

Human-centered computing [**Accessibility**]: Accessibility technologies; Applied computing [**Life and medical sciences**]: Health informatics

Keywords

Trouble-indicating behavior, dementia, classification

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive neuro-degenerative disease that deteriorates memory (short- and long-term), executive capacity, visual-spacial reasoning, and linguistic ability [3]. Caregivers who assist individuals with AD at home are common, but their involvement is often the precursor to long-term care [5]. As populations age, the incidence of AD will double or triple, with Medicare costs alone reaching $189 billion in the US by 2015 [1]. Given the growing need to support this population, there is an increasing interest in the design and development of technologies that support this population at home and extend one's quality of life and autonomy.

ASSETS'14, October 20–22, 2014, Rochester, NY, USA.
ACM 978-1-4503-2720-6/14/10.
http://dx.doi.org/10.1145/2661334.2661382.

We are designing intelligent dialog software that can engage in two-way communication for two purposes: a) to help guide individuals towards the completion of daily household tasks, and b) to fulfill social functions. Our goal is to encode in software the techniques used by caregivers to help their patients achieve these activities; this includes automatically identifying and recovering from breakdowns in communication. Here, we consider conversational data between patients and interviewers and develop methods of feature analysis and classification to identify confusion.

Trouble indicating behaviors (TIBs) are indications that the speaker requires aid to resolve phonological, morphological/syntactic, semantic, or discourse confusion [10]. There are 12 TIBs: **1)** neutral or non-specific requests for repetition (local), **2)** request for confirmation – repetition with reduction, **3)** request for confirmation – complete repetition, **4)** request for confirmation – repetition with elaboration, **5)** request for specific information, **6)** request for more information, **7)** correction of semantic inaccuracy, **8)** lack of uptake / lack of continuation, **9)** hypothesis formation (guessing), **10)** metalinguistic comment, e.g., *I can't remember.*, **11)** reprise / minimal dysfluency, e.g., *Eerrr, I want to – we went to the river.* , **12)** request for repetition – global, e.g., *wait – go back to the part about....*

2. EXPERIMENTS

We use the Carolina Conversations Collection (CCC) [7] and DementiaBank [2]. The CCC consists of conversations between older adults (> 60 years) and young adult interviewers. There are 31 interviewees diagnosed with AD (7 male) and 41 interviewees without AD (9 male). DementiaBank is a longitudinal collection of conversations where 196 older adults with dementia and 98 matched controls performed the 'cookie theft' picture description task with an interviewer [6] annually. Audio and textual transcriptions, annotated temporally at the utterance level, are available in each case. Here, every utterance is further annotated with TIBs by a speech-language pathologist and members of the research team.

We extract over 200 lexical/syntactic and acoustic features from these data, which cannot all be enumerated here. Instead, we use a method similar to [4] except instead of a t-test criterion, we use an analysis of variance (ANOVA) to rank features according to how well they separate classes, according to

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\sum_i n_i(\bar{x_i} - \bar{x})^2/(K-1)}{\sum_{ij}(x_{ij} - \bar{x_i})^2/(N-K)},$$

Table 1: The means (μ) and variances (σ^2) of the top 5 most relevant features to TIB discrimination, according to the ANOVA method.

Feature	No TIB μ (σ)	TIB μ (σ)
Brown bigram model (negative log-likelihood)	-11.35 (4.15)	-10.735 (2.27)
Words/minute	177.31 (1480.10)	164.47 (2244.89)
Ratio of PRP to NN+PRP	0.40 (0.127)	0.487 (0.137)
Mean 2nd MFCC	-2.62 (3.53)	-3.14 (3.33)
% of strong neutral words	0.136 (0.0731)	0.0794 (0.041)

Table 2: Accuracy of TIB identification across databases and classifiers, given either all features or the top 15 as determined by the ANOVA method.

Classifier	Features	Database	
		CCC	DementiaBank
NB	Top 15	79.5%	67.0%
	All	63.1%	63.1%
SVM	Top 15	71.0%	59.2%
	All	55.7%	68.4%
Adaboost	Top 15	48.3%	65.0%
	All	26.7%	58.3%

where $K = 2$ since we are comparing TIB to non-TIB utterances, and N is total number of samples. Table 1 shows the top five most relevant features according to this method, in order: 1) the log-likelihood of the utterance given a bigram model trained with MLE on the Brown corpus; 2) words/minute; 3) the ratio of pronouns (PRP) to nouns (NN) and PRP, given the Stanford tagger [9]; 4) the mean of the 2nd Mel-frequency cepstral coefficient over the utterance; 5) the proportion of words that are strong neutral, obtained from the MPQA subjectivity lexicon [11].

We compare three binary classifiers that differentiate TIB utterances from non-TIB utterances in table 2. Specifically, we use naïve Bayes (NB) to model the likelihood of an utterance given the class with a Gaussian using maximum *a posteriori* training, a support vector machine (SVM) with the Gaussian kernel, and Adaboost, which is a common ensemble method based on iteratively building a number of decision trees to focus training on the relatively 'difficult' examples [8]. In general, results are promising except for Adaboost in the CCC database.

3. DISCUSSION

This paper presents an analysis of lexical/syntactic and acoustic features that are indicative of *trouble indicating behaviors* in the speech of individuals with AD across two popular databases. Although some of these features are expected (e.g., confusion tends to be related to slower rates of speech), others are more surprising (e.g., MFCC features generally refer to the transfer function of the vocal tract). With feature selection and simple naïve Bayes classification, up to 79.5% of utterances can be correctly classified as having TIB or not. Future work will include refinement of classification methods and integration of this functionality into automated personal assistants for use by people with AD.

4. ACKNOWLEDGMENTS

This work is funded by the Toronto Rehabilitation Institute - University Health Network and the Natural Sciences and Engineering Research Council of Canada (RGPIN 435874).

5. REFERENCES

[1] A. J. Bharucha, V. Anand, J. Forlizzi, M. A. Dew, C. F. R. III, S. Stevens, and H. Wactlar. Intelligent assistive technology applications to dementia care: Current capabilities, limitations, and future challenges. *American Journal of Geriatric Psychiatry*, 17(2):88–104, February 2009.

[2] F. Boller and J. Becker. *DementiaBank Database Guide*. University of Pittsburgh, 2005.

[3] J. L. Cummings. Alzheimer's disease. *New England Journal of Medicine*, 351(1):56–67, 2004.

[4] K. Fraser, F. Rudzicz, and E. Rochon. Using text and acoustic features to diagnose progressive aphasia and its subtypes. In *Proceedings of Interspeech 2013*, Lyon France, August 2013.

[5] J. E. Gaugler, F. Yu, K. Krichbaum, and J. Wyman. Predictors of nursing home admission for persons with dementia. *Medical Care*, 47(2):191–198, 2009.

[6] H. Goodglass and E. Kaplan. *The Assessment of Aphasia and Related Disorders*. Philadelphia, PA: Lea and Febiger, 2nd edition, 1983.

[7] C. Pope and B. H. Davis. Finding a balance: The carolinas conversation collection. *Corpus Linguistics and Linguistic Theory*, 7(1):143–161, 2011.

[8] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. In *Machine Learning*, pages 80–91, 1999.

[9] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 173–180. Association for Computational Linguistics, 2003.

[10] C. M. Watson. An analysis of trouble and repair in the natural conversations of people with dementia of the Alzheimer's type. *Aphasiology*, 13(3):195 – 218, 1999.

[11] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005.