UNC
DEPARTMENT OF COMPUTER SCIENCE

Detecting Linguistic Characteristics of Alzheimer's Dementia by Interpreting Neural Models

Sweta Karlekar, Tong Niu, and Mohit Bansal

Introduction

Alzheimer's disease (AD) is an **irreversible and progressive brain disease** that can be stopped or slowed down with medical treatment. However, current medical evaluation techniques are lengthy and include:

- Mental status and mood testing
- Physical and neurological exams
- Surveying extensive medical history
- Brain imaging

Language changes serve as a sign that a patient's cognitive functions have been impacted. Automated detection can potentially lead to a **speedy and early diagnosis**, while manual inspection of language requires significant time and expertise.

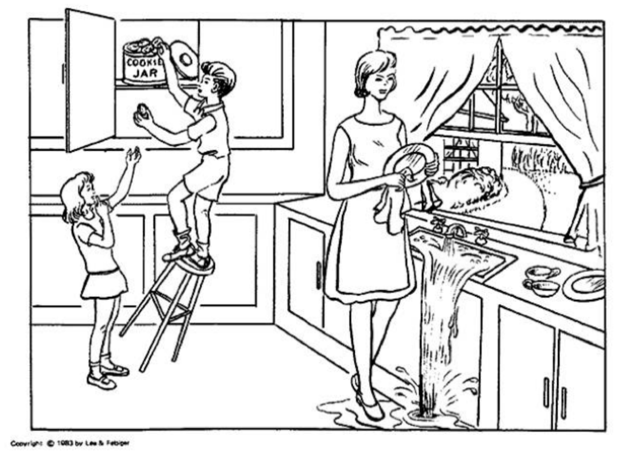
Task:
Use NLP techniques to **classify and visualize/analyze/interpret** the linguistic characteristics of Alzheimer's patients using transcripts of spoken language.

- Apply **three neural models** based on CNNs, LSTM-RNNs, and their CNN-RNN combination, to distinguish between language samples from AD and control patients
- Achieve a **new independent benchmark accuracy** for the AD classification task
- Provide analysis based on **activation clustering and first-derivative saliency** techniques
- Analyze **gender-separated** AD data

Methods & Dataset

This study utilizes DementiaBank (Boller and Becker, 2005), the largest publicly available dataset of transcripts and audio recordings of AD (and control) patient interviews.

Patients are asked to perform multiple tasks, such as being given an image and asked to describe what they see. See the figure on the right for the image given to patients for the Cookie Task.



Each transcript comes with:

- Standard part-of-speech tagging
- Description of tense
- Repetition markers

Within the 14362 utterance samples, 11458 come from transcripts of Alzheimer's-diagnosed interviewees and 2904 from those of control patients (majority baseline AD-positive=79.8%).

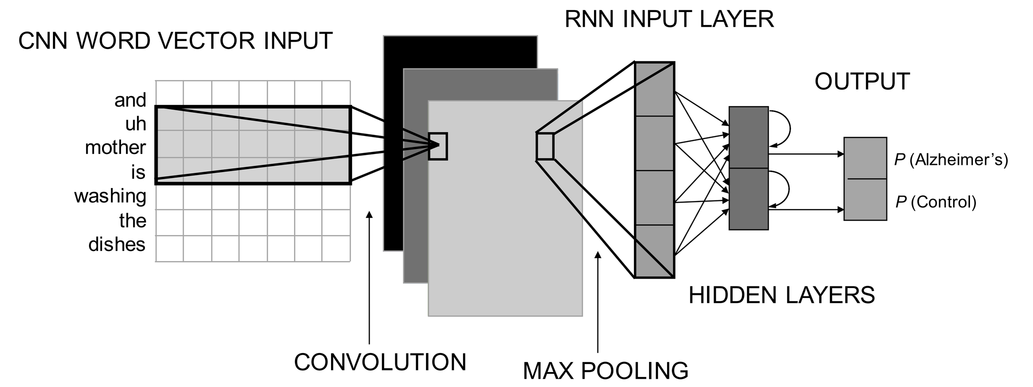
Patients are given multiple description tasks and recorded → Transcripts and recordings are made available on DementiaBank → Automatic morphosyntactic speech tagging is performed → Individual utterances are split to be used as data points → Data is presented in two setups: with and without POS tags

Models

CNN: Embedding and a convolutional layers are applied, followed by a max-pooling layer. The convolution features are obtained by applying filters of varying window sizes to each window of words. The result is then passed to a softmax layer that outputs probabilities over two classes (i.e. AD-positive and AD-negative).

LSTM-RNN: Our LSTM-RNN model consists of an embedding layer followed by an LSTM layer. The final state, containing information from the entire sentence, is fed to a fully-connected layer followed by a softmax layer to obtain the output probabilities.

CNN-LSTM: Laying an LSTM-RNN layer on top of a CNN layer allowed us to combine both models' complementary strengths:



Results

Model	Details	Accuracy
2D-CNN	Non-Tagged Utterances	82.8
LSTM	Non-Tagged Utterances	83.7
CNN-RNN	Non-Tagged Utterances	84.9
CNN-RNN	POS-tagged Utterances	**91.1**

With untagged data, our CNN, LSTM and CNN-LSTM models achieved an accuracy of 82.8%, 83.7%, and 84.9%, respectively. When fed with the given POS-tagged data, our best-performing CNN-LSTM model achieved 91.1% accuracy, setting a new benchmark accuracy for this task (Orimaye et al., 2017, 2016, 2015; König et al., 2015; Rudzicz et al., 2014).

Analysis

Gender Uniformity: We found that the sets of the top ten most common POS-tags for both AD-positive men and women are the same.

Moreover, our best performing model achieved 86.6% classification accuracy on solely the male data and 86.2% accuracy on solely the female data, demonstrating that it found **no statistically significant difference between the AD-positive languages of men and women**.

Activation Clusters: Activation clustering treats the activation values of n neurons per input as coordinates in an n-dimensional space (Girshick et al., 2014; Aubakirova and Bansal, 2016). Inputs that maximally activate similar neurons are clustered using K-means.

Rediscovering Existing Strategies: Our activation clusters corroborated previously known linguistic characteristics of Alzheimer's disease (Watson, 1999; Rudzicz et al., 2014).

- Short answers and bursts of speech:
  {'okay', 'and', 'yes', 'oh!', 'yes', 'fine'}
- Repeated requests for clarification:
  {'Did I say facts?', 'Did I get any?', 'Did I say elephant?'}
- Starting with interjections:
  {'Well I gotta see it', 'Oh I just see a lot of uh...', 'So all the words that you can...'}

Novel Automatic Pattern Discovery: The most common POS tags in each cluster are found automatically to allow us to better understand which grammatical structures are favored.

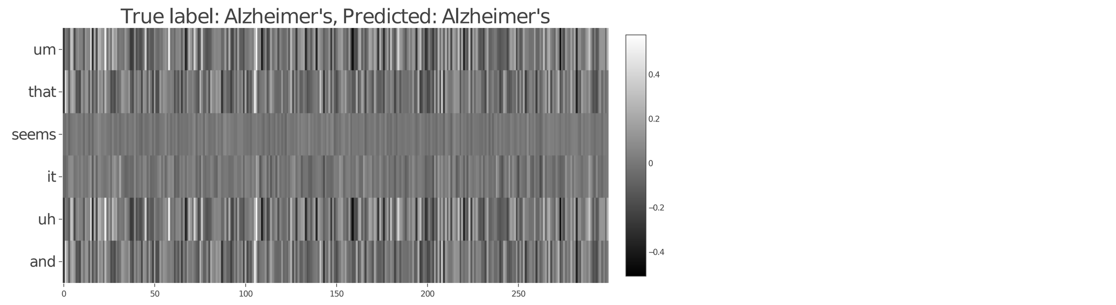
For the same task, non-AD clusters contain the same top four POS tags. Qualitative analysis shows dissimilarities in the most common POS tags between the Cookie task's AD and non-AD cluster(s). The AD-positive cluster has only two most-used POS tags in common with the non-AD cluster. In fact, none of the 3 non-AD clusters have adjectives or adverbs in their most-used POS tags list, unlike the AD-positive cluster.

Non-AD Clusters - Cookie Task					
POS	Freq	POS	Freq	POS	Freq
n	0.15	*n*	0.13	*n*	0.15
det	0.13	*det*	0.13	*det*	0.13
presp	0.07	*part*	0.09	*part*	0.10
part	0.05	*presp*	0.09	*presp*	0.10

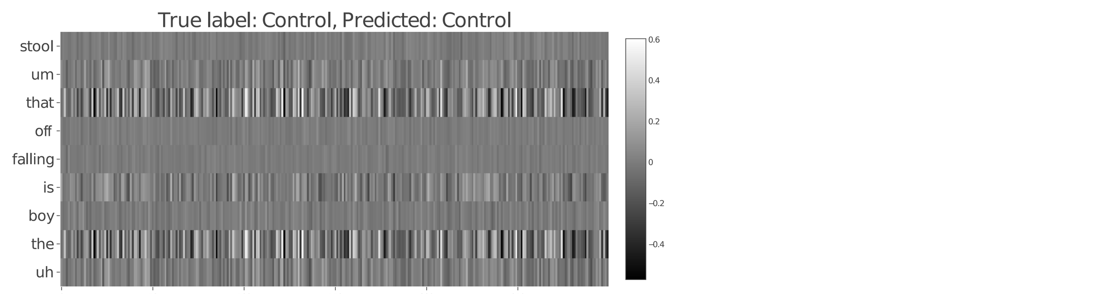
AD		Non-AD	
POS	Frequency	POS	Frequency
n	0.20	*n*	0.15
det	0.14	*det*	0.13
adj	0.05	*presp*	0.07
adv	0.04	*part*	0.05

v = verb, *n* = noun, *pro* = pronoun, *adv* = adverb, *det* = determiner, *aux* = auxiliary verb, *prep* = preposition, *co* = interjection, *part* = participle, *presp* = present participle

First Derivative Saliency Heat Maps: Saliency heat maps illustrate which words in an input had the biggest impact on the classification of the whole sentence (Simonyan et al., 2013). This is done by taking the gradient of the final score with respect to the word embeddings of the inputs.

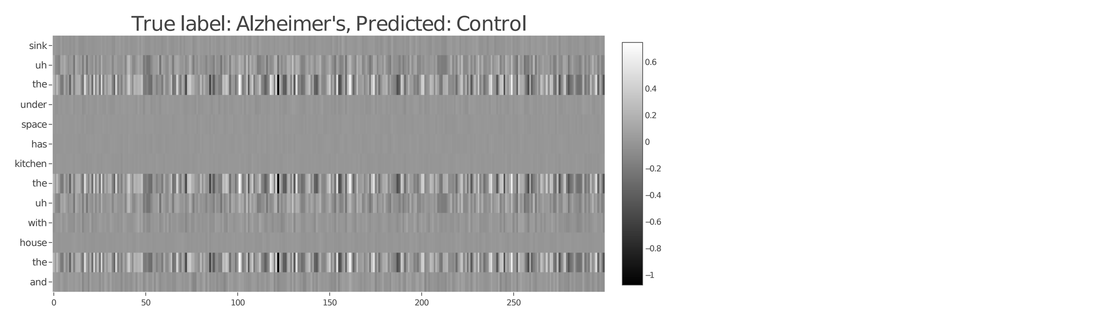


The filler words "uh", "um", and the initial "and" are emphasized, showing that they have a lot of influence on the classification of this example as AD-positive.



The "uh" filler word is not highlighted, showing that most control patients do not use filler words as heavily as Alzheimer's patients. Definite articles and determiners that give specificity or structure to a sentence have the biggest impact on classification (e.g., "the", "that").

Visualizing Model Limitations with Saliency Heat Maps



The model misclassified this example due to the longer length of the utterance and the heavy use of determiners. However, the repeated "uh"s and starting with a coordinating conjunction instead indicate AD, which confused our model. More contextual data from the transcript and advanced neural methods are needed to help the model better classify samples that exhibit both AD and control features.

Conclusion

We applied three models to the AD classification task, and our CNN-LSTM model achieved a **new benchmark accuracy** in classifying AD using neural models. We illustrated with two visualization techniques how these models capture unique linguistic features present in AD patients. We also discussed gender analysis and found no statistically significant difference in the AD-positive language of men and women.

Potential future work includes using more conversational context and implementing multi-class classification to differentiate among stages of AD. We also plan to generalize this model to other neurological diseases, such as Diffuse Lewy Body disease and Huntington's disease.

References

- Malika Aubakirova and Mohit Bansal. 2016. Interpreting neural networks to improve politeness comprehension. In Proceedings of EMNLP.
- Francois Boller and James Becker. 2005. Dementiabank database guide. University of Pittsburgh.
- Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587.
- Alexandra König, Aharon Satt, Alexander Sorin, Ron Hoory, Orith Toledo-Ronen, Alexandre Derreumaux, Valeria Manera, Frans Verhey, Pauline Aalten, Phillipe H Robert, et al. 2015. Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1(1):112–124.
- Sylvester O Orimaye, Jojo SM Wong, Karen J Golden, Chee P Wong, and Ireneous N Soyiri. 2017. Predicting probable alzheimers disease using linguistic deficits and biomarkers. BMC bioinformatics, 18(1):34.
- Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong, and Judyanne Sharmini Gilbert Fernandez. 2016. Deep-deep neural network language models for predicting mild cognitive impairment. In BAI@ IJCAI, pages 14–20.
- Sylvester Olubolu Orimaye, Kah Yee Tai, Jojo SzeMeng Wong, and Chee Piau Wong. 2015. Learning linguistic biomarkers for predicting mild cognitive impairment using compound skip-grams. arXiv preprint arXiv:1511.02436.
- Frank Rudzicz, Leila Chan Currie, Andrew Danks, Tejas Mehta, and Shunan Zhao. 2014. Automatically identifying trouble-indicating speech behaviors in alzheimer's disease. In Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility, pages 241–242. ACM.
- Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.
- Caroline M Watson. 1999. An analysis of trouble and repair in the natural conversations of people with dementia of the Alzheimer's type. *Aphasiology*, 13(3):195–218.

Acknowledgements
This work was supported by a Google Faculty Research Award, a Bloomberg Data Science Research Grant, an IBM Faculty Award, and NVidia GPU awards.