Dementia Detection Through Topic and Thought Process Analysis

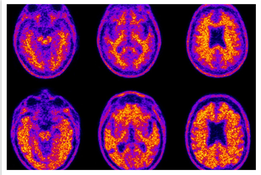
Noemi Andras, Jessica Borowy, Soyoon Lee, Mariya Pasheva, & Ivana Pavlovic

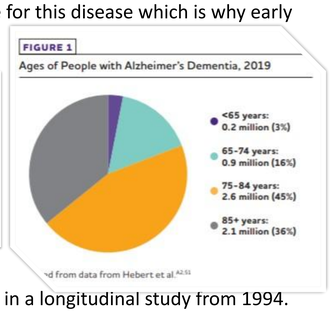
Advisers: Natalie Parde, Taha Khan, Joseph Hummel, Renata Revelo

BACKGROUND

Dementia

- According to the Merriam-Webster dictionary, dementia is a “usually progressive condition marked by the development of multiple cognitive deficits.”
- On average it impacts 50 million individuals worldwide
- Difficult to diagnose due to the diagnosis process being time consuming and costly. Diagnosis for early stages is not reliable.
- There is no known cure for this disease which is why early detection is vital.





DementiaBank Dataset

- This data was collected in a longitudinal study from 1994.
- The data consisted of audio transcripts of interviews with patients who were diagnosed with dementia and those who were healthy.
- Patients were asked to describe as many details from an image known as the Cookie Theft Image, which is often used in psychological research related to dementia.

PROPOSED SOLUTION

- Our proposed solution is to consider the Cookie-Theft Image as a “safe” zone and to determine when patients are wandering away from the “safe” zone in terms of their descriptions of the image.
- We will create new features for the dementia detection software, that was developed by Professor Natalie Parde and her graduate students, that will help increase the software's accuracy when detecting if a patient has dementia based on how much and how far they stray from the topics displayed in the image.
- This solution was inspired by a previous study that developed an app called iWander. The app determines when dementia patients wandered outside of a specified safe zone and tried to help them find their way back into the safe zone through auditory alerts and instructions.

HYPOTHESIS

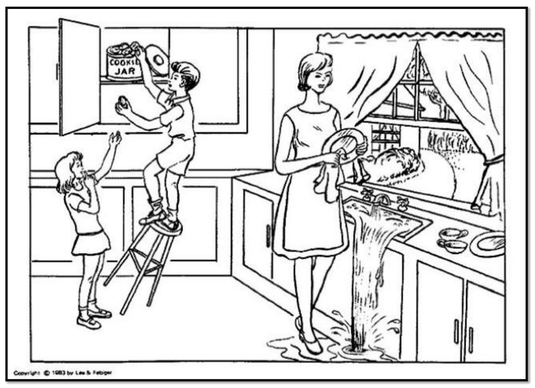
Our hypothesis is that dementia patients engage in more ‘conversational wandering’ than healthy controls, sporadically jumping from one area of the image to another when describing it. This hypothesis is supported by psycholinguistic evidence suggesting that dementia patients tend to go off-topic and engage in verbal repetition when providing narratives.

METHODOLOGY

I. Learn about the NLTK Python library in order to determine which tools will be most useful for the required analysis of the interview transcripts.
II. Analyze and annotate randomly selected files from the dataset manually in order to determine what information will be most vital in determining the location of which portion of the picture the patient is currently describing and to extract topics and inferences that patients make about the image. Then use a vector of binary features to automate the process for the remaining files.

Dialog Act Type	Description	Example
Pronouns	Parts of Speech that describe the subject of the sentence, identify what it refers to	"She", "He", "They", "I", "You", "It"
Weather (a type of inference)	Anything that has to do with seasons or outdoor conditions	"Summer", "gentle breeze" , "warm"
Objects	Non human nouns	"Cookie Jar", "Sink"
Relationships (a type of inference)	Interpretations of the connections between the people in the picture	"Brother", "Sister", "Mother", "Twins"
Inferences/Asumptions	Whenever the participant describes things that are not explicitly in the picture (weather can be included in this or we can separate them into categories later)	"There is green grass outside","It must be", "Perhaps", "Seems to be", "Looks like..."
People/Age Group	gender of person as well as indication of how old they are	"child, adult, toddler...etc"
Actions	human or object movements	"running, washing, etc."
Location	space where objects or people exist	"kitchen, house, outdoors, garden"
Clothes	Anything the people in the picture have on their person	"dress, tennis shoes, etc"
Interactions	between who/what with what	"boy standing on stool"

III. Isolate a small section of the dataset to be used to test the new features.
IV. Divide the image into zones based off the topics depicted in the image.
V. Train a logistic regression classification model after development has been completed.
VI. Test the model with the small portion of the dataset that was set aside earlier in the process.
VII. Conduct error analysis for specific topics and sections in the image
VIII. Compile final report based off the results found and determine if the model's overall accuracy has increased.



EXPECTED RESULTS

- We expect that our vector of binary features will be able to accurately record whenever a patient mentions a specific topic in the Cookie Theft Image, and it will be able to record the proper zone ID of where in the image that topic can be found.
- Using this binary vector we expect to be able to follow the path the patient took when describing the picture and, in the process, determine what their thought process was.

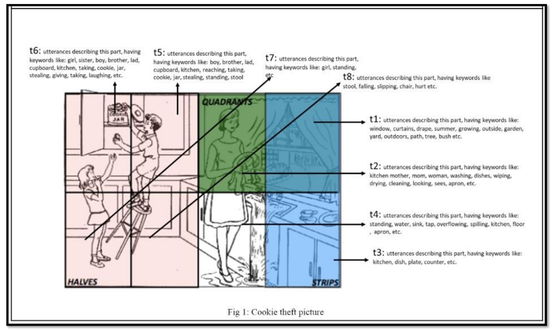
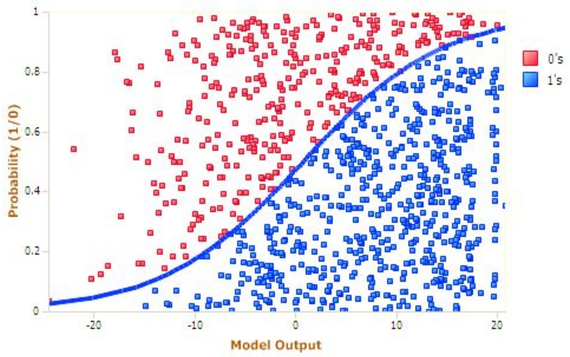


Fig 1: Cookie theft picture

- We would then use an image divided similarly like the one above in order to depict which zones the patient “traveled” to in order to visually represent whether they wandered off topic during their description.
- 80% of the data obtained by the binary vector would then be used to train a logistic regression classification model which would determine if a patient had dementia or not.
- 20% of the data obtained by the binary vector would be used to test the model to determine how accurately it could tell the difference between a dementia patient and a healthy one.



- We expect that our model would be able to distinguish between the two groups with 97% accuracy or better. If some patients are incorrectly classified, we would use those patients to calculate the experimental error and retrain the model to hopefully decrease the amount of misclassifications.

PyCharm

ACKNOWLEDGEMENTS

We would like to acknowledge the following people:

- Professor Natalie Parde and her graduate students: Flavio, Mina, and Shahla for allowing us to work on their software, helping us obtain the necessary data that we needed, and answering all our questions related to the project.
- Our graduate TA Taha Khan for helping us to determine our next steps in the project and aiding us whenever we were facing an issue during the research process.
- Professor Joseph Hummel and Professor Renata Revelo for giving us the opportunity to participate in research and for offering support and advice whenever needed.

REFERENCES

Becker JT, Boiler F, Lopez OL, Saxton J, McGonigle KL. The Natural History of Alzheimer's Disease: Description of Study Cohort and Accuracy of Diagnosis. *Arch Neurol.* 1994;51(6):585–594. doi:10.1001/archneur.1994.00540180063015

Di Palo, F., & Parde, N. (2019). *Enriching Neural Models with Targeted Features for Dementia Detection* (pp. 1–7). Chicago, IL: University of Illinois.

Alzheimer’s Association, (2019). *2019 Alzheimer's Disease Facts and Figures*. Retrieved March 3, 2020, from https://www.alz.org/alzheimers-dementia/facts-figures