Dementia Detection Through Topic and Thought Process Analysis

Noemi Andras, Jessica Borowy. Soyoon Lee, Ivana Pavlovic
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607
nandra4@uic.edu, jborow7@uic.edu, slee651@uic.edu, ipavlo2@uic.edu

1 Introduction

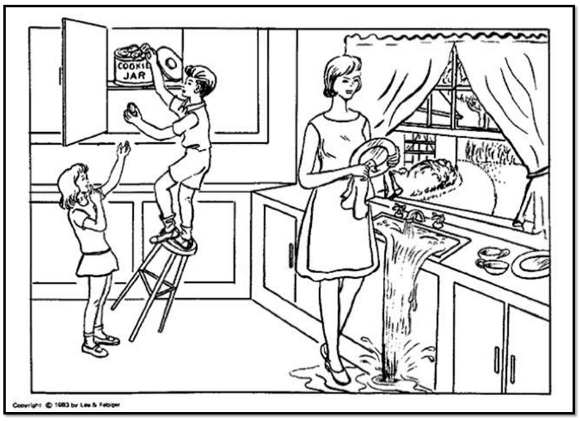
Dementia has been severely detrimental to the population by being one of the "leading causes of death in the world and [affecting] at least 50 million individuals" [6]. Since the disease currently has no cure it is imperative to be able to detect it as early as possible. One of the ways to detect dementia early in its progression is by considering speech patterns of patients (for example, their use of filler words, repetitions, incomplete words). That is why natural language processing (NLP) tools can provide potential guidance for health professionals as well as possible detection of dementia among patients. This research examines the differences in conversational patterns between dementia patients and age-matched healthy controls completing the Cookie Theft Picture Description Task in the DementiaBank dataset, focusing specifically on the differences between (a) the topics discussed, and (b) the order in which they are mentioned.

2 Related Work

Currently, there is no handy or reliable way to detect dementia in its early stages [6]. Possible signs of patients having dementia include the use of filler words, repetitions, incomplete words and a slower rate of speech [1]. Due to these symptoms, speech pattern analysis is useful for dementia detection, making natural language processing (NLP) a viable choice for this field of research. Furthermore, NLP has also been shown to be "[effective when] processing clinical notes" [8].

Previous research has looked at the publicly available DementiaBank dataset from the University of Pittsburgh School of Medicine [7]. The dataset focuses on transcripts of patient interviews regarding an image known as the Cookie Theft picture shown in Figure 1. This image is useful because it contains a variety of different details and requires the patient to have the cognitive ability to accurately describe them. Since the dataset distinguishes patients with dementia from healthy patients, this allows researchers to gain insight into speech patterns of dementia patients.

Figure 1: Cookie Theft Image

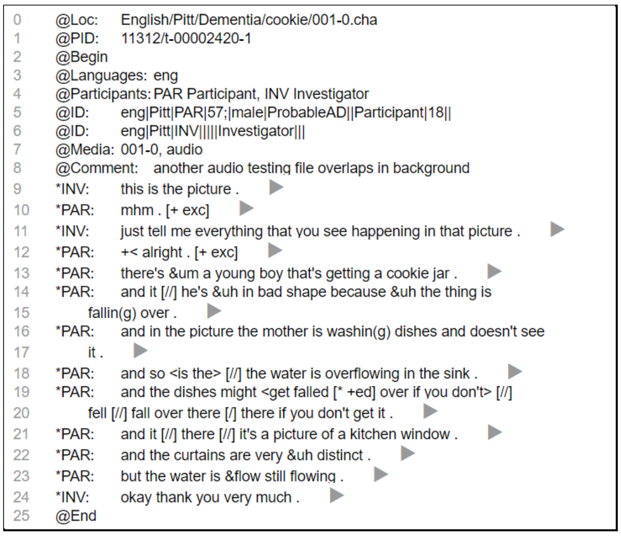


Using the DementiaBank dataset, researchers in the field of NLP have been able to create dementia detecting frameworks resulting in high success rates. Previous research [4] addressed the imbalance of the DementiaBank dataset by doing class weight correction. This way, they achieved high accuracy (88%) and precision (93%) that outperformed a previous model developed by Zhou et al. [9]. As of now, the highest cross-validation accuracy on the DementiaBank dataset is 97%, achieved using an attention-based hybrid network [6].

3 Data Model

The data model that was utilized during this research project was the aforementioned DementiaBank dataset from a longitudinal study conducted in 1994 by the University of Pittsburgh School of Medicine. The patient transcripts from the DementiaBank dataset were saved as .cha files in different folders (one for the control group and one for the dementia group) and could be listened to or viewed in a readable format like the one shown in Figure 2. The files contained many special characters related to non-verbal responses from the patient such as: long pauses, breaths, laughs, and others. The files contained audio responses from both the interviewer and the patient as well as a file ID and a designation for dementia or control patients. These transcripts also varied in length as some patients had more to say than others.

Figure 2: Example Patient Transcript



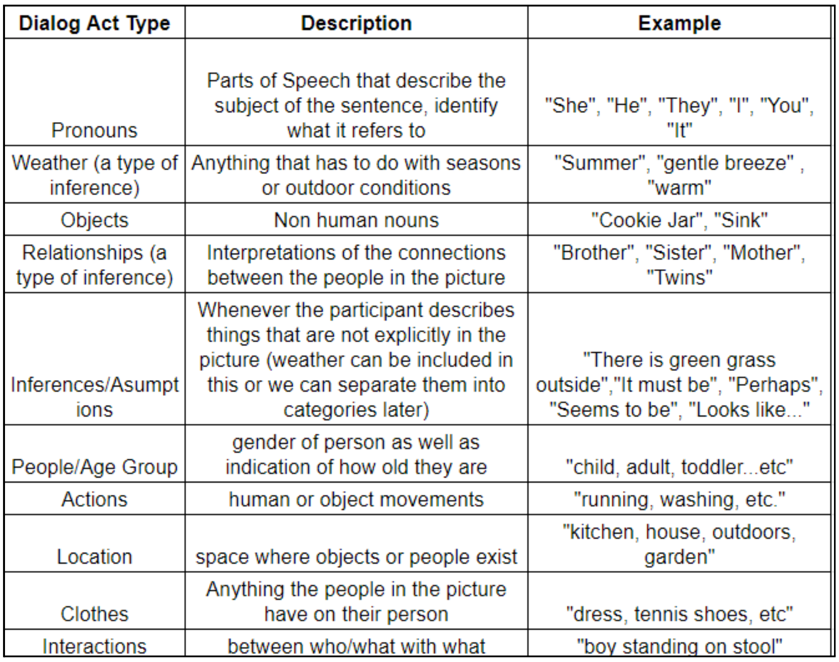
```
@Loc:     English/Pitt/Dementia/cookie/001-0.cha
@PID:     11312/t-00002420-1
@Begin
@Languages: eng
@Participants: PAR Participant, INV Investigator
@ID:      eng|Pitt|PAR|57;|male|ProbableAD||Participant|18||
@ID:      eng|Pitt|INV|||||Investigator|||
@Media: 001-0, audio
@Comment:   another audio testing file overlaps in background
*INV:     this is the picture .
*PAR:     mhm . [+ exc]
*INV:     just tell me everything that you see happening in that picture .
*PAR:     +< alright . [+ exc]
*PAR:     there's &um a young boy that's getting a cookie jar .
*PAR:     and it [//] he's &uh in bad shape because &uh the thing is
     fallin(g) over .
*PAR:     and in the picture the mother is washin(g) dishes and doesn't see
     it .
*PAR:     and so <is the> [//] the water is overflowing in the sink .
*PAR:     and the dishes might <get falled [* +ed] over if you don't> [//]
     fell [//] fall over there [/] there if you don't get it .
*PAR:     and it [//] there [//] it's a picture of a kitchen window .
*PAR:     and the curtains are very &uh distinct .
*PAR:     but the water is &flow still flowing .
*INV:     okay thank you very much .
@End
```

4 Experimental Setup

Initially, before feature design and model training could begin, we had to learn about the NLTK Python library in order to determine which tools would be most useful for the required analysis of the interview transcripts. This included learning how to properly tokenize each file so that only the investigator responses were removed as well as ensuring that the characters related to non-verbal patient responses were left in each file. The designation of whether the given transcript was from the dementia group or from the control group also had to be removed in order to ensure that the model did not base its classification decisions on those designations.

We then had to 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. We developed annotation guidelines shown in Figure 3 so that each annotation is as unbiased as possible. These manual annotations would then be used in the future to create a vector of binary features to automate the process for the remaining files.

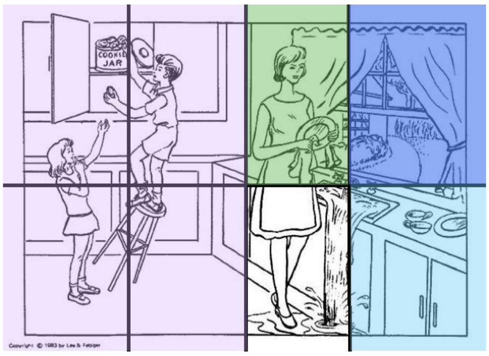
Figure 3: Annotation Guidelines Used During Manual Annotations



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"

This vector of binary features would include a set of columns that would contain the specific content that we were looking for throughout the file and enter a 0, if that content was not found in the file and a 1 if the content was present. Each row of the vector would correspond to one particular patient transcript and would contain the ID of that file in order to be able to determine which files were already featurized and which ones still required feature extraction. This file ID could also be used by the researchers during the training and testing of the classification model in order to determine whether or not the model was correctly labeling each file with dementia or control. Finally, the vector would record what sector of the image each feature corresponded to by recording the sector IDs of the Cookie Theft Image. These sectors and their subsequent IDs would be designated by the researchers as depicted in Figure 4.

Figure 4: Example of Cookie Theft Image sectors



Approximately 80% of the data instances would be used to train a logistic regression classification model while the remaining 20% would be used for testing. After the completion of testing, an analysis of the classifier's predictions would be used to conduct error analysis for patient classifications.

5 Evaluation and Expected Results

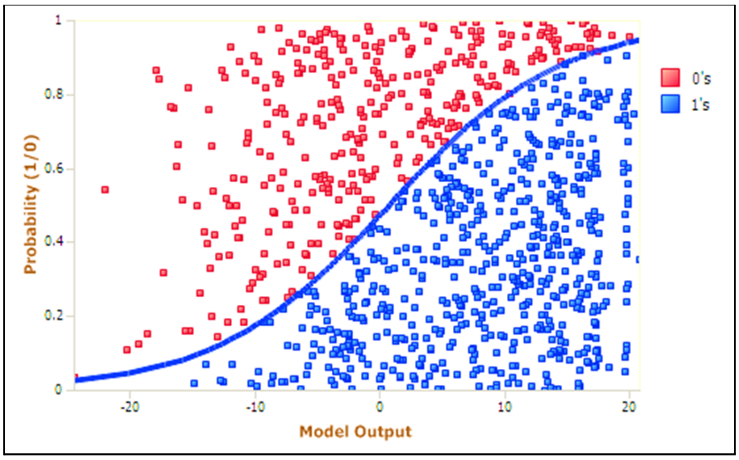
The Cookie-Theft Image will be considered as a "safe" zone that will determine when patients are wandering away from it in terms of their descriptions of the image. The goal is to understand the differences between thought processes of healthy and dementia diagnosed patients in order to apply this knowledge in future detection of dementia. In doing so, we hope to improve upon the accuracy of prior work in automated dementia detection [4].

We expect to see 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 [3].

In future research, the new features would be evaluated using a logistic regression classification model. This model is what would be used to test if the developed features were accurately determining which patients had dementia and which were healthy. The model would sort patients into categories depending on the 'path' they took through the image with their descriptions and create a graph like the one shown in Figure 5. In the chart, the different colored data points represent the two different patient groups, whereas the line dividing them represents the model's accuracy in classifying each data point. This model would also help determine the percentage of misclassifications that occurred during the experiment and would be used to determine what further training the new features require.

Figure 5 is what is expected for the model to behave like, and we hope to have results like that. In order to achieve that, the algorithm calculates probabilities, considers a specific datapoint and determines the probability of it belonging to a specific group or classification. The probability we would want this model to calculate for our experiment would be: P(Patient has dementia | a large amount of 'conversational wandering'), read as: the probability that a patient has dementia given that they experienced a large amount of wandering throughout the image as they described it. The logistic regression model calculates this probability in the terms of 'pass' or 'fail' and only becomes a classifier once certain conditions provided by the programmer are provided (i.e. if passed then the current patient has dementia, if failed then the current patient is healthy).

Figure 5: Logistic Regression Classification Model



We expect that our model would be able to distinguish between the two groups with 97% accuracy or better.

6 Discussion and Conclusion

Over the course of this project we were exposed to many new topics and new Python libraries. We were able to explore the field of Natural Language Processing (NLP) which none of us had been previously exposed to. We discovered the many different applications of NLP such as automated translations, speech to text recognizers, spell checking, and many more. Arguably the most important lesson that we learned from this research was about the time and effort that goes into designing an NLP software. There are many things that need to be determined before you even begin coding such as, the types of words and languages that you expect your software to be able to recognize and work with, the way that you want a person's speech to be interpreted and what should be done with this interpretation, and how you will train and test your model to ensure that the software performs with the highest accuracy.

We also got to learn about how to tokenize files using the NLTK library in Python to help us get the important information out of each of the patient transcripts. We explored the Spacy Named Entity Recognizer tool which we were thinking of using to help automate our annotation process for each transcript since it could read through the file and assign a descriptive label to each word in the file. These labels would let us know if the word was referring to a person, object, date, etc. However, while working on the project we discovered that the named entity recognizer was incorrectly labeling a lot of the words in the file (for example, it considered many things to be works of art which were not present in the transcripts at all). We attempted to train the named entity recognizer to use labels that were similar to our feature categories in the hopes that it would learn to correctly label the words in the way we wanted. Unfortunately, this did not go according to plan and we were forced to try a different method of automating this process. Eventually, we settled on attempting to create a binary feature vector.

Another topic we were able to explore was machine learning. We were exposed to this through our research on the type of classification algorithm we were going to use to test our 'wandering' program's accuracy. The logistic regression classification model is a commonly used machine learning classification algorithm that is typically used when a software is sorting data points into different categories. A well known example of this is a software that is designed to determine the difference between a picture that contains a cat and one that contains a dog. After training the model with many such pictures the programmer would use a logistic regression classification model to determine how well the computer could actually tell the difference between the two pictures and which pictures were misclassified. We also briefly explored other classification models such as Decision Trees, Random Forest, and Naive Bayes.

All in all, we are very grateful to have had the opportunity to work on this project and to learn things that we would not have otherwise. We would like to thank our advisor, Professor Parde, for supervising the project and directing us in the right direction. We would also like to thank Taha Khan, the teacher assistant in the Early Research Scholars Program (ERSP) for his guidance and help during the semester. We will be continuing this project in the future and look forward to learning even more aspects of NLP and machine learning that we can use in any of our future endeavors.

References

[1] C. I. Guinn and A. Habash, “Language analysis of speakers with dementia of the Alzheimer’s type,” presented at the 2012 AAAI Fall Symposium Series, 2012.

[2] D. Beltrami, G. Gagliardi, R. R. Favretti, E. Ghidoni, F. Tamburini, and L. Calza, “Speech Analysis by Natural Language Processing Techniques: A Possible Tool for Very Early Detection of Cognitive Decline?,” vol. 10, p. 369, Nov. 2018.

[3] E. Reeve, P. Molin, A.Hui and K. Rockwood, “Exploration of verbal repetition in people with dementia using an online symptom-tracking tool,” Int Psychogeriatr. 2017 Jun

[4] F. Di Palo and N. Parde, “Enriching Neural Models with Targeted Features for Dementia Detection,” 2019.

[5] F. Sposaro, J. Danielson, and G. Tyson, “iWander: An Android application for dementia patients.” pp. 3875–3878, 2010.

[6] J. Chen, J. Zhu, and J. Ye, “An Attention-Based Hybrid Network for Automatic Detection of Alzheimer’s Disease from Narrative Speech.” 2019.

[7] S. O. Orimaye, J. S.-M. Wong, and C. P. Wong, “Deep language space neural network for classifying mild cognitive impairment and Alzheimer-type dementia,” vol. 13, no. 11, p. e0205636, 2018.

[8] X. Zhou, Y. Wang, S. Sohn, T. M. Therneau, H. Liu, and D. S. Knopman, “Automatic extraction and assessment of lifestyle exposures for Alzheimer’s disease using natural language processing,” vol. 130, p. UNSP 103943, Oct. 2019.

[9] Chunting Zhou,Chonglin Sun,Zhiyuan Liu,and Fran-cis Chi-Moon Lau.2015. A c-lstm neural network for text classification.CoRR,abs/1511.08630.