Literature Review

Jomar Alcantara

2018-07-13

Contents

1 Introduction

Alzheimer's Disease (AD) is a neurodegenerative disease in which the brain develops neurofibrillary tangles and neuritic plaques along with the loss of cortical neurons and synapses. The hallmark clinical symptoms are cognitive deficits such as problems with episodic and semantic memory, organising and planning, difficulties with language and visuospatial deficits[1]. In addition, these symptoms are often accompanied by emotional difficulties such as depression and irritability and behavioural difficulties. Whilst there are several variants of Dementia, AD remains the most common type of Dementia and will be the focus of this review.

According to a recent report commissioned by the Alzheimer's Society in 2015, they estimate the prevalence of AD in the UK at approximately 815,000 people. This represents 1 in 14 of the population aged 65 or over and 1 in 79 in the general population. They estimate an annual healthcare spend on 4.3 billion of which approximately 85 million is spend solely on diagnosis and that the total impact of AD (excluding the costs associated with early onset dementia) is 26.3 billion annually. Globally, this picture is a lot bleaker. A recent report suggests that in 2015 there were 46 million people with a diagnosis of AD and that number is expected to hit 131.5 million by 2050 [2]. The report also states that the worldwide cost of AD in 2018 is estimated to be in the region of one trillon US dollars.

Current thinking suggests that the cognitive deficits associated with AD often begin before the clinical symptoms of the disease become apparent. Researchers propose that neurofibrillary tangles and other associated physiological effects of AD develop over time until a threshold is reached and clinical symptoms become apparent [3]. If this is the case, it should be possible to detect subtle cognitive changes in language and memory before a clinical diagnosis can be formed. Given that language is less intrusive to test and requires a lot of the cognitive processes that may be impacted by AD, a lot of research has focused on the decline in the use of language in those with AD.

One of the most famous pieces of research was by Berisha and Liss (2015) [4] who examined speeches and public interviews of former US president Ronald Reagan. They found that Reagan's speeches towards the end of his presidency suffered from difficulties in word-finding, inappropriate phrases and uncorrected sentences which are hallmarks of language deterioration associated with Alzheimer's Disease. Another classical study by Snowdon et al (1996) [5] looked at whether linguistic ability in early life was asoociated with cognitive function and AD in later life. They found that idea density (defined as the number of expressed propositions divided by the number of words) was a key predictor in predicting whether nuns would go on to develop AD in later life. They found that those who would go on to develop AD all had low idea density in early life and they found no AD present in those with high idea density.

The range of language deficits in those who suffer with AD are wide and varied and differ as the disease develops. According to the DSM 5 [1], those with Mild dementia suffer from noticeable word finding difficulty. They may

substitute general terms for more specific terms and may avoid the use of specific names of acquaintances. There may be grammatical errors involving subtle omission or incorrect use of articles, prepositions, auxiliary verbs, etc. Those who have progressed from Mild to Major depression also have difficulties with expressive or receptive language. They will often use general-use phrases such as "that thing" and "you know what I mean" and prefers general pronouns rather than names. With severe impairment, sufferers may not even recall names of closer friends and family. Idiosyncratic word usage, grammatical errors, and spontaneity of output and economy of utterances occur. Echolalia (meaningless repetition of another person's spoken words) and automatic speech typically precede mutism. With the wide range of deficits someone with AD can suffer, it makes sense to try to categorise these deficits in some way.

Emery [6] completed a literature review in this area and she looked at four levels of language: Phonology, Morphology, Syntax and Semantics. Her review looked at the idea that language and the processes involved in language are hierarchical. She proposes that language goes from simple units of construction, and build layers of complexity and sophistication. She found that people with AD generally had intact Phonology and Morphology but more impaired Syntax and Semantics. Emery concludes that language decline is hierarchical and is related to the complexity of the language task given to a participant. She and that language decline is hierarchical in that the language forms we learn last (the most complex language forms) are the first to deteriorate.

It is clear from both the clinical diagnostic criteria and supporting research that language is impacted in those with AD. However, one of the costs of analysing language is that there is a huge burden on trained practitioners, be it clinical psychologists, audio transcribers and text encoders to facilitate the process of collecting data and analysis. The field of machine learning and natural language processing has been suggested as a way toimprove the accuracy and lessen the human cost of this research as well as provide new insights into the difficulties that AD suffer in terms of language decline.

The purpose of this review is to seek to understand what developments have been made in the use of Machine Learning and Natural Language processing to aid the diagnosis of Mild Cognitive Impairment (MCI) and AD. A search of the literature was conducted using ProQuest (PsychArticles), SCOPUS, Web of Science. The following results were found (Table 1). All papers were then reviewed for relevance by reading the abstract and full text where appropriate and a shortlist was compiled. An additonal search through references of shortlisted papers was also conducted. Papers were included where there was an attempt to classify participants as MCI, AD or Healthy using speech and/or language by using Machine Learning techniques. We excluded any papers that focused on other forms of dementia or cognitive impairment, as well as any papers in which the language being analysed was not English. This resulted in 21 journal articles and/or conference papers which form this review.

Database	Number of Results	Search Terms
ProQuest(PsychArticles)	1484 Results	Language AND Decline AND Dementia
ProQuest(PsychArticles)	486 Results	Language AND Decline AND Dementia AND Speech
ProQuest(PsychArticles)	159 Results	Machine Learning AND Dementia AND Language
Web of Science	1207 Results	Language AND Decline AND Dementia
Web of Science	151 Results	Language AND Decline AND Dementia AND Speech
Web of Science	34 Results	Machine Learning AND Dementia AND Language
Scopus	791 Results	Language AND Decline AND Dementia
Scopus	91 Results	Language AND Decline AND Dementia AND Speech
Scopus	29 Results	Machine Learning AND Dementia AND Langauge

2 Types of Language Assessment

Test

2.1 Picture Description Tasks

One of the most commonly used tasks to measure lexical analysis is the Picture Description task. In dementia research, the most common task picture description task is the Cookie Theft picture. This does very well in giving a valid dataset of language but suffers somewhat from ecological validity. In essence, when completing this task, the participant is generating speech to describe the picture but does not use some of the skills generally used in every day interactions. If the goal is to a way

In assessing whether a client has AD, part of a clinicans job is to assess language capability. As described above, AD has a significant impact on language and a clinician has access to tools designed to test lanaguage capabilities. An example of this is part of the Boston Diagnostic Aphasia Examination (BDAE), called the Boston Cookie Theft picture description task [7] . In this task participants are asked to describe a picture presented to them in as much detail as possible. The picture itself depicts a familiar domestic scene and would not require participants to use any vocabulary beyond that learned in childhood. It was originally designed to assess Aphasia, but has shown itself to be useful in the diagnosis of AD as well [?]

2.2 Other types of task

3 Connected Language and potential features

Connected Language and it's study can be defined as self generated discourse. This is important as this requires a number of cognitive processes which we know are impacted in those who suffer from dementia. One of the common problems in the analysis of connected language is the ability to differentiate between what is a purely language related problem vs a memory problem.

3.1   (

n-grams) One of the first features discussed as a potential predictor of MCI or AD is the n-gram. An n-gram is a contiguous sequence of n items from a given sample of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. For example, given the sequence of words "to be or not to be", this extract is said to contain six 1-gram sequences (to, be, or, not, to, be), five 2-gram sequences (to be, be or, or not, not to, to be), four 3-gram sequences(to be or, be or not, or not to, not to be) and so on. This is useful as, given a large portion of text or speech, we can predict the probability of a word being close by to a given word. One of the first attempts to use machine learning and natural language techniques to look was conducted by Thomas [8] who was able to successfully demonstrate the ability of machine learning algorithms to analyse n-grams as well as other features to outperform a naive rule-based classifier which always selects the modal class.

Orimaye et al (2017) [9] investigated the use of machine learning algorithms to detect differences in syntactic, lexical and n-gram linguistic biomarkers to distinguish between those with probably AD and healthy controls, The authors found significant differences in the uses of all three types of biomarkers for those with AD and healthy controls. Their results show that the top 1000 n-gram features plus the twenty-three syntactic and lexical features was the most successful at differentiating the two groups (0.93 AUC)). However, there were some limitations of the study. The context of the audio content was very specific, limited to one specific description task, and therefore using general speech as proposed in this project would potentially use a different set of linguistic features which may or may not have similar predictive power. In addition, the transcriptions were encoded using the CHAT format which is a framework for manually annotating speech and the challenge would be finding a way to automate this. This paper is noteworthy as it provides evidence that using machine learning to identify participants with MCI is possible and it provides some ideas of potential features which can be used. Unique to this paper is the use of n-gram features, which could be explored in the proposed project.

Asgari, Kaye and Dodge (2017) also looked at the linguistic characteristics of older adults with mild cognitive impairment (MCI) vs healthy controls where they hypothesised that they would be able to predict those with MCI, a distinguishing characteristic of Alzheimer's disease and other variations of dementia. Using recordings of unstructured conversations (with standardized preselected topics across subjects) between interviewers and interviewees they grouped spoken words using Linguistic Inquiry and Word Count (LIWC) which is a technique used to categorize words into features such as negative and positive words. They then applied support vector machines (SVM's) and random forest classification algorithms to investigate whether machine learning could be used to distinguish between those with MCI and healthy controls. They were able to successfully used machine learning algorithms to distinguish between these two groups with an accuracy of 84. The authors report that this method is highly reliant on high-fidelity transcription of the conversations which is labour intensive, but they anticipate that technology is advancing sufficiently quickly that automated high-quality transcriptions are possible in the near future. This

paper provides a different perspective in how to tackle the problem deriving language features from speech using a different framework. The challenges both Orimaye et al and these authors experienced were around the automation of transcription and this would potentially be an area to explore in the proposed project. Asgari et al (2017) used the Linguistic Inquiry and Word Count (LIWC) methodology, developed by Pennebaker et al (2007) to generate features for their paper. Pennebaker, Boyd, Jordan and Blackburn (2015) have since developed their framework. The LIWC is designed to categorize and evaluate the various emotional, cognitive and structural components which are present within samples of speech and/or written text. This method, Linquistic Inquiry and Word Count (LIWC) has evolved from it's initial incarnation in 1993 to it's last update in 2015. The premise is the use of words from particular categories provide an insight into the psychological processes and/or diagnoses an individual has. The latest version of this methodology uses a dictionary of 6400 words, word stems and select emoticons and assigns these constructs to various categories. This has been updated to include modern uses of language such as 'text speak'. This manual goes summarises the process of constructing the framework as well as discussing the reliability and validity of the measure and a review of studies using the LIWC was conducted by Tausczik and Pennebaker (2010) support the notion that the LIWC is valid across multiple psychological domains. This paper is useful as it provides a useful and validated framework from which the proposed project can potentially derive language features.

Currently formal diagnosis of depression is expensive and time-consuming involving both the use of questionnaires and the use of a trained professional to assess an individual. Schwartz et al, explored the use of language as an aid to diagnose depression in a naturalistic setting i.e. facebook status updates. The authors used both the LIWC (described above) and n-grams as features to distinguish those with depression from healthy individuals and were able to track participants levels' of depression through language successfully.Whilst the authors were able to use a reasonably sound method of measure degree of depression and have evidenced a methodology that is able to track a levels of depression over time, it would be more helpful to use a more clinically relevant measure of depression and anxiety i.e., PHQ-9 (Kroenke, 2001) and GAD-7 (Spitzer et al, 2006) to add validity to their findings. This paper is interesting as it uses both LIWC and n-grams as a method of deriving features from text / speech for depression which, as described above, have also been used as features to distinguish MCI from healthy participants. Given the comorbity of depression and dementia (Meyers, 1998), it would be interesting for the proposed project to further analyse the link between dementia and clinical depression via the use of language. This paper also supports the idea that language changes over time and is a marker for deterioration in both those with depression and dementia.

Fraser, Meltzer and Rudzicz (2015) looked at connected speech using the DementiaBank corpus. They found that there were four factors which informed the classification of participants as either healthy or AD. These four factors were semantic impairment, acoustic abnormality, syntactic impairment and information impairment.

Lopez-de-Ipina (2018) have since moved on

4 Conclusion

Identification of language impairment is important in Dementia because it aids diagnosis of specific types of dementia, which in turn can alter the prognosis and change the management of the degenerative disorder. As these differences in language are quite subtle, the varying subtypes of dementia are frequently misdiagnosed and can sometimes be missed altogether.

Given the burden on the diagnosis of dementia on clinicians, it appears to be useful to find some non-invasive protocols for the early diagnosis of dementia. It has already been shown that analysis of speech and language has shown markers that pre-date the official diagnosis of dementia (Snowdon et al, 1996;)[4]. A significant amount of research has gone into the use of machine learning techniques to potentially look at the automated classification of participants with MCI and/or AD, however this is a new area of research and there are some gaps in our knowledge. Firstly, the vast majority of the research described above looks at pre-existing datasets like the DementiaBank corpus (CITE). Whilst this is useful for 'backtesting the data', it would be useful for this process to be tested live to see whether these results can be replicated now.

Future research should be directed towards developing non-intrusive ways of detecting subtle changes in language based in the home, such that any deterioration that could indicate the presence of MCI or AD and be flagged up for further review via referral. Given the amount of processing power and the sophistication of machine learning, in particular deep learning algorithms. Further, despite the quality of datasets being used to 'backtest' these algorithms, further research should look at generating further datasets to increase the validity of the results found so far as well as using other methods to generate data other than Picture Description tasks which we could claim are limited in scope. It would be worth looking at how we can utilse these technologies further to aid detect the earliest signs of cognitive impairment.

References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5, 2013.

[2] Martin Prince, Anders Wimo, M Guerchet, Gemma-Claire Ali, Yu-Tzu Wu, and Matthew Prina. World Alzheimer Report 2015 The Global Impact of Dementia. Technical report, 2015.

[3] Peter J Nestor, Philip Scheltens, and John R Hodges. Advances in the early detection of Alzheimer's disease. *Nature Medicine*, 12(8):961–966, 2006.

[4] Visar Berisha, Shuai Wang, Amy LaCross, and Julie Liss. Tracking Discourse Complexity Preceding Alzheimer's Disease Diagnosis: A Case Study Comparing the Press Conferences of Presidents Ronald Reagan and George Herbert Walker Bush. *Journal of Alzheimer's Disease*, 45(3):959–963, 2015.

[5] David A. Snowdon, Susan J. Kemper, James A. Mortimer, Lydia H. Greiner, David R. Wekstein, and William R. Markesbery. Linguistic ability in early

life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association*, 275(7):528–532, 1996.

[6] V O Emery. Language impairment in dementia of the Alzheimer type: a hierarchical decline? *International Journal of Psychiatry in Medicine*, 30(2):145–164, 2000.

[7] E. Kaplan, H. Goodglass, and S. Weintraub. Boston Naming Test. In *The Corsini encyclopedia of psychology*, page 2009. 2010.

[8] C. Thomas, V. Keselj, N. Cercone, K. Rockwood, and E. Asp. Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. *IEEE International Conference Mechatronics and Automation, 2005*, 3(February 2005):1569–1574, 2005.

[9] Sylvester O. Orimaye, Jojo S.M. Wong, Karen J. Golden, Chee P. Wong, and Ireneous N. Soyiri. Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinformatics*, 18(1):1–13, 2017.