Computational cognitive assessment: investigating the use of an Intelligent Virtual Agent for the detection of early signs of dementia

Bahman Mirheidari[1], Daniel Blackburn[2], Ronan O'Malley[2], Traci Walker[3], Annalena Venneri[4], Markus Reuber[5], and Heidi Christensen[1,6]

[1]Department of Computer Science, [2]Sheffield Institute for Translational Neuroscience, [3]Department of Human Communication Sciences, [4]Department of Neuroscience, [5]Academic Neurology Unit, Royal Hallamshire Hospital, [6]Centre for Assistive Technology and Connected Healthcare, University of Sheffield, UK

Introduction

Dementia

- can affect a person's cognitive abilities, memory, ***speech*** and ***language***.
- the number of people developing dementia is increasing, and early diagnosis (using ***automatic***, ***accurate*** and ***low-cost*** tools) is ***desired***.
- qualitative methodology of ***Conversation Analysis (CA)*** can identify ***communication problems*** of people talking with neurologists [1, 2], but it is ***expensive*** and ***difficult*** to ***scale up*** for routine clinical use.
- we have developed an ***Intelligent Virtual Agent (IVA)*** [3, 4] who asks a series of ***memory-probing*** questions, mimicking the style of questions used during the ***history taking*** part of a normal ***face-to-face consultation***.

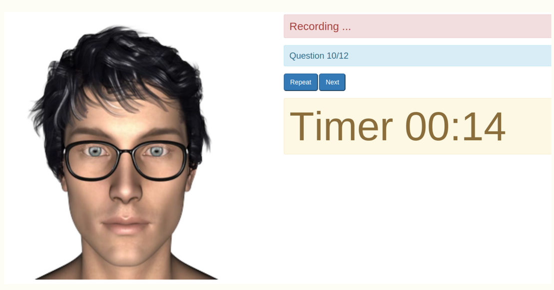

Figure 1: IVA

Research focus

- ***feasibility*** of eliciting conversations with people with memory problems, i.e. the IVA acts as a neurologist or a ***Digital Doctor***.
- ***applicability*** of using an IVA in the diagnostic pathway by augmenting the initial conversation-based assessment to include more standard test procedures, such as administering ***verbal fluency tests***.
- ***expanding*** our diagnostic categories to include healthy elderly controls, Mild Cognitive Impairment (MCI), as well as Functional Memory Disorder (FMD), and Neurodegenerative Dementia (ND) to reflect the ***variety of conditions*** seen in practice.

Dementia detection system

- Diarisation (***who talks when***) (Kaldi diarisation toolkit), ASR (Kaldi toolkit), Feature extraction (NLTK python + Praat toolkit), Classifier (Linear Regression from Scikit-learn python), IVA (https://www.botlibre.com;pre-recorded human voice)

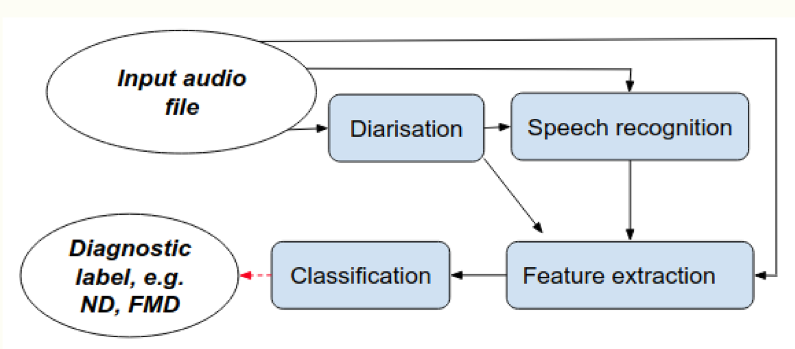

Figure 2: Block diagram of dementia detection system

Experiment

Data sets

hallam :interviews between neurologists and patients in Hallamshire hospital; **seizure** :interviews between neurologists and people with epilepsy (Seizure); **iva** :conversations between the IVA and the subjects.

Table 1: Data set info including Length: the total length in hours/mins, Utts.: number of utterances, Spks.: number of speakers, and Avg. Utts.: Average utterance length in seconds.

Data set (No)	Length	Utts.	Spks.	Avg. Utts.
HALLAM (45)	12h	8,970	117	4.8s
SEIZURE (241)	50h 16m	28,000	597	6.3s
IVA (61)	4h 20m	1,944	85	8.0s

IVA participants

- Functional Memory Disorder (**FMD**):10; Neurodegenerative Dementia (**ND**):19; Mild Cognitive Impairment (**MCI**):18; Healthy Control (**HC**):14

Extracted features

- **CA-inspired:** 14 features (Pat: Patient) + 3 for APs (Accompanying Person(s)), e.g. patient answered me for who's most concerned question, average number of empty words.
- **Acoustic-only:** 12 features (Pat) + 12 for APs, e.g. silence, intonation, pitch, H1-H2.
- **Lexical-only (Part of Speech):** 12 features (Pat) + 12 for APs, e.g. number of verbs, nouns adverbs.
- **Word Vector:** 7 features after Principal Component Analysis (PCA).
- **Verbal fluency test:** (naming animals (semantic fluency test)/words begin with the letter 'P' (phonemic fluency test) in one minute) 6 features, number of unique animals/P-words correctly uttered, average and standard deviation of Age of Acquisition (AoA).

Results

A. Diarisation and speech recognition

The HALLAM and SEIZURE data sets were used for training the i-vector based diarisation module (using the Probabilistic Linear Discriminant Analysis (PLDA) and the Bidirectional Long Short Term Memory/Time-Delay Neural Network (BLSTM)-TDNN based ASR.

- totally **unseen data**
- for the 18 recordings of the IVA with manual transcripts, the Diarisation Error Rate (DER): **11%**, and the Word Error Rate (WER): **59%**.

B. Classification accuracy k=10 fold cross validation

Methodology: comparing conversation-only to fluency test-only and then to the combination of conversation and fluency test.

Table 2: Classification accuracy. Conv.: Conversational question, Flu.: Verbal fluency test, *:the most significant features.

Features (No)	FMD/ND/MCI/HC	FMD/ND	FMD/MCI	ND/MCI	FMD/HC	ND/HC	MCI/HC
Conv.(72)	43%	**79%**	68%	51%	54%	**88%**	**78%**
Flu.(6)	39%	69%	**71%**	51%	50%	73%	69%
Conv.+Flu.(78)	**48%**	**79%**	**71%**	**57%**	**58%**	85%	72%
Conv.+Flu.(22)*	***62%***	***79%***	***75%***	***68%***	***63%***	***94%***	***88%***

C. Feature selection

C. Feature selection Recursive Feature Elimination (RFE) [5]

Table 3: The most significant (22) features.

Rank	Feature	Feature type
1	ApsAvgSil	Acoustic-only
2	PatAVPauses	CA-inspired
3	PatAvgSil	Acoustic-only
4	PatSemSTDAoA	Fluency semantic
5	ApsAVUniqueWords	CA-inspired
6	APsAVTurnLength	CA-inspired
7	PatAVFillers	CA-inspired
8	PatSemCount	Fluency semantic
9	WV_col5	Word vector
10	PatPhnAVGAoA	Fluency phonemic
11	PatSemAVGAoA	Fluency semantic
12	WV_col4	Word vector
13	WV_col7	Word vector
14	APsNoOfTurns	CA-inspired
15	WV_col1	Word vector
16	WV_col3	Word vector
17	PatFailureExampleEmptyWords	CA-inspired
18	PatAVUniqueWords	CA-inspired
19	WV_col2	Word vector
20	PatAVTurnLength	CA-inspired
21	PatAVAllWords	CA-inspired
22	PatNoOfTurns	CA-inspired

D. Receiver Operating Characteristic curve

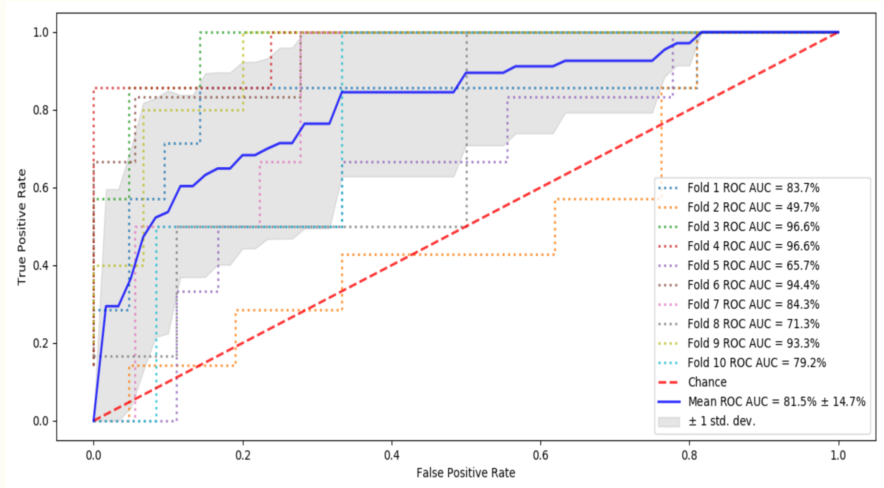

Figure 3: ROC-AUC (the most significant features).

Conclusions and further work

Conclusions

- we explored the feasibility and applicability of using the IVA to administer standard dementia screening tests.
- we extracted a variety of features.
- adding the fluency test improves accuracy for the 4-way classification achieving 62%.
- applying the feature selection the ND/HC was the easiest binary classification (94%), while the ND/MCI was the hardest (68%).

Future work

- Expanding to include ***more types of feature***
- ***Improving*** the ASR, diarisation and feature extraction modules.
- ***Improving the IVA*** to make it more responsive.

Acknowledgements

This research has been supported by the MRC Confidence in Concept Scheme and the EU H2020 Marie Skłodowska-Curie programme TAPAS (Training Network for Pathological Speech processing; Grant Agreement No. 766287)

References

[1] Elsey, C., Drew, P., Jones, D., Blackburn, D., Wakefield, S., Harkness, K., Venneri, A. & Reuber, M.(2015). Towards diagnostic conversational proles of patients presenting with dementia or functional memory disorders to memory clinics. Patient Education and Counseling, 98, 1071-1077.

[2] Jones, D.,Drew, P., Elsey, C., Blackburn, D., Wakefield, S. ,Harkness, K., & Reuber, M. (2015). Conversational assessment in memory clinic encounters: interactional profiling for differentiating dementia from functional memory disorders. Aging & Mental Health, 7863, 1-10.

[3] Mirheidari, B., Blackburn, D., Harkness, K., Walker, T., Venneri, A., Reuber, M. &Christensen, H. (2017). An avatar-based system for identifying individuals likely to develop dementia. Proc. Interspeech, 3147-3151.

[4] B. Mirheidari, D. Blackburn, A. Venneri, M. Reuber,T. Walker, & H. Christensen, H. (2018). Detecting signs of dementia using word vector representations. Proc. Interspeech.

[5] Pedregosa, F. & Varoquaux, G. (2011). Scikit-learn: Machine learning in python. Journal of Machine Learning Research, 12, 2825-2830.