An investigation into diagnosing Alzheimer’s disease type and progress by applying NLP techniques to a combination of Interactional and Linguistic features.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**[1/4]**

Front Page -1 page

Disclaimer 1 page

Acknowledgements 1 page

Content list 1 page

Abstract ½ - 1 page

**[2/4] - (6-8 pages)**

**Introduction**

Project Objective ½ page

Background:

What is Alzheimer’s? ½ page

<https://www.google.com/patents/US6475161>

How is Alzheimer’s currently diagnosed? ½ page

<https://www.google.com/patents/US6475161>

How is Alzheimer’s currently diagnosed using Linguistics? ½ page

Literature review: What have other done?1.5 - 2 pages

- Fraser Narrative Linguistics - Uses Cookie Data

-Comparative Study of Oral and Written Picture Description in Patients with Alzheimer’s Disease - Uses Cookie Data

-Effects of working memory deficits on the communicative functioning of Alzheimer’s dementia patients

Motivation - criticize other people’s work ½ page

There has been no study done on Interaction Features using computational techniques.

Why use interactional features?

Where did not using IF not perform as well?

Where did not using IF not perform as well?

Define your interactional features.

Include studies on IF’s - fillers, pauses, etc

**[3/4] Methods & Materials: - What have you done? (24 pages)**

Requirements: Model. (½ page)

Model design: (7 pages)

Feature plan - based *loosely* off Fraser top 30. Plan was to add in IF’s to investigate improvement.

Describe what higher level groupings that Fraser used. -Did not use acoustic features.

Feature extraction:

List of NIF’s - Fraser (~30) 2

Your version of each features.

Include references and paper for “Info Units” etc

List of IF’s (37) - 2-3 pages

Hypothesis and theory behind each code up of each feature

Dataset (2-3 pages)

Talk Bank - Dementia Bank

Implementation: (3 pages)

Python Used (NLTK)

PreProcessing Transcripts (CHAT/CLAN)

Regex used

Applied stanford parser - Higher accuracy rate and statistically sounder when compared against Fraser

Wary of using Transcript data - Fine for NIF, not so fine for IF

Evaluation methods: (6 pages)

Used NIF as baseline

Different labelings of target

Investigated DT, LR, KNN, NB models

How you came to build result table:

Took top NIF features

Selected top using kbest - Correlation charts

Evaluated P/R/F1

Took top IF features

Selected top using kbest

Evaluated P/R/F1

Merged 2 Tops and reran to find if any IF features performed better than any NIF

Hypothesis (see email): My reasoning is, that if the new top 22 include any IF, then that means a Non-IF has been "bumped" from the kbest selection. (½ page)

Discussion/Conclusion (½ page)

The final 22 features resulted in being made up of 40% IF (Count: 9) and 60% Non-IF (Count: 13) so one can argue that IF's can be used to improve the classification of the diagnosis.

Critical Analysis (2 pages)

Talk through results table

Talk through why some features improve the classifier more than others.

Honest appraisal (2 pages)

Why LR worked so well.

Think it’s a poor predictive model but when just investigating IF vs NIF it works well.

My NIF’s probably not as robust as Fraser

Further work (Prepare before viva - Big question! ) What would you do if you had time (6 pages)

STIR - more work on repair - Similarity comparison of spoken and intended word

Rerun through DA

Analysis of other non-Cookie data

Need to find proxy controls for it

[4/4] (1-2 pages) - At LEAST 10 refs, ideally 20

References/Bibliography

Appendices

More detailed material that is not crucial to understanding of main message

**FORMAT - AIM FOR 50 PAGES INCLUDING APPEN**

MAX 60 (75 pages with appen)

MIN 38 (without appen)

Deadline - Aug 23rd Final report, code (zipped), presentation slides

Campus vivas - Tues 26th Aug - Fri 8th Sept (Maybe week of 11th)

**VIVAS**

Presentation - 10-15 mins (10 slides ish, 15 max)

Demonstration - if appropriate, 5-10 mins

Q and A - 10 mins

Aims:

Explain and justify your work

Demonstrate what you have achieved

Impress and interest examiners

Show ability to respond to related questions.

Bring laptop as back up - slides ready to go