Similarity Measures for the Detection of Clinical Conditions with Verbal Fluency Tasks

Felipe Paula, Rodrigo Wilkens, Marco Idiart and Aline Villavicencio
Introduction
Verbal Fluency

- Semantic Verbal Fluency test which is used in neuropsychological assessment

- Semantic Verbal Fluency
  - Category (e.g. animal)
  - As many words possible in a limited time
  - No repeated words

Patient

dog, cat, cow, horse, giraffe, elephant
Verbal Fluency

Patient

Chain

Switch

dog, cat, cow, horse, giraffe, elephant
Verbal Fluency

(Examples of chain categories)

Pets: dog, cat, parrot, hamster, fish, ...

Farm animals: cow, horse, sheep, chicken, ...

Africa’s animals: giraffe, elephant, lion, zebra, ...

Cats: lion, tiger, cat, panther, jaguar, ...
Objectives

Investigate similarity measures for detecting switches in word sequences

Describe a method using switch information to predict clinical conditions
Our work
**Data**

- **Fluency Data**  \( n = 100 \)

Bertola et al, 2014, *The Parsing System ‘Palavras’: Automatic Grammatical Analysis*, 2000 Aarhus University Press.
Materials

- **WordNet**
  Knowledge-based

- **GloVe**
  Word Embedding

- **Pointwise Mutual Information**
  Association measure

- **Corpus**
  PT-BR wikipedia
  44,345 types and 118,095,637 tokens
  Lemmatized with PALAVRAS (Bick 2000)

Eckhard Bick, *The Parsing System 'Palavras': Automatic Grammatical Analysis*, 2000 Aarhus University Press.
WordNet

Source: NLTK book, https://www.nltk.org/book/ch02.html
Methods

GloVe

Zebra

Lion

Cat

Dog

Word

\[ n_1, n_2, n_3, n_4, n_5, n_6, n_7, \ldots, n_n \]
Methods

Pointwise Mutual Information

\[
\text{pmi}(x; y) \equiv \log \frac{p(x, y)}{p(x)p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}.
\]
Goal: find the switches in a test

Transform the VFT in a vector

Apply a function that indicate if a pair is a switch or not
## Heuristics

| Pairwise Word Similarity | Dog  | Cat  | Cow  | Horse | Zebra | Monkey | Fish  |
|--------------------------|------|------|------|-------|-------|--------|-------|
|                          | 0.81 | 0.10 | 0.75 | 0.67  | 0.61  | 0.32   |       |
**Global Mean Detection**

Decision made using the mean of the sequence of word pairs similarity.
Heuristics

Local Mean Detection

Decision made using the mean of the last $k$ pairs' similarity
Voting Detection

Decision made by voting between local and global detections
Experiments

Feature Extraction

- Fraction of Smallest Chain
- Average Chain Length
- Largest Chain Size
- Number of Switches

Classifier
Random Forest
Experiments

Verbal Fluency Data
- ALZ
- aMCD
- aMCD
- Healthy

Judge Switch Identification
- Healthy
- X
- Cognitive Impairment
- AD + aMCD + mMCD

Automatic Switch Identification
- Healthy
- X
- a+mdMCI

GloVe Cosine Similarity

PMI Association Strength

WordNet Path Similarity

Healthy X aMCI

Healthy X AD
Results
Results

Which heuristics presented best performance?

Numeric best classifier (AUC)

CI

- Vote3 PMI: 0.76
- Gold Standard: 0.68

aMCI

- Local1 WN: 0.68
- Gold Standard: 0.58

a+mdMCI

- Local1 GloVe: 0.75
- Gold Standard: 0.67

AD

- Vote3 PMI: 0.93
- Gold Standard: 0.82
Results

Which heuristics presented best performance?

Winning combinations

**aMCI**

- $V_{1\_wn}$: 0.63
- $V_{2\_wn}$: 0.64
- $V_{3\_pmi}$: 0.54

**a+mdMCI**

- $V_{1\_wn}$: 0.7
- $V_{2\_wn}$: 0.7
- $V_{3\_pmi}$: 0.73

**AD**

- $V_{1\_wn}$: 0.87
- $V_{2\_wn}$: 0.89
- $V_{3\_pmi}$: 0.93
Conclusions
Discussion and Future Work

Our simple heuristic switch detection based classifier seem to outperform judge based classifier.

This can be valuable tool in the detection of clinical conditions

Voting strategies specially using WordNet and PMI work well to predict the groups.

Future works include the study of this methodology in other clinical populations and the relationship with another neuropsychological tests.
Thanks

Felipe Paula (Institute of Informatics - UFRGS)
  felipesfpaula@gmail.com

Rodrigo Wilkens (CENTAL - Université Catholique de Louvain)
  rswilkens@gmail.com

Marco Idiart (Institute of Physics - UFRGS)
  marco@gmail.com

Aline Villavicencio (Institute of Informatics - UFRGS)
  alinev@gmail.com