

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**

dog, cat, cow, horse, giraffe, elephant

Patient

Verbal Fluency

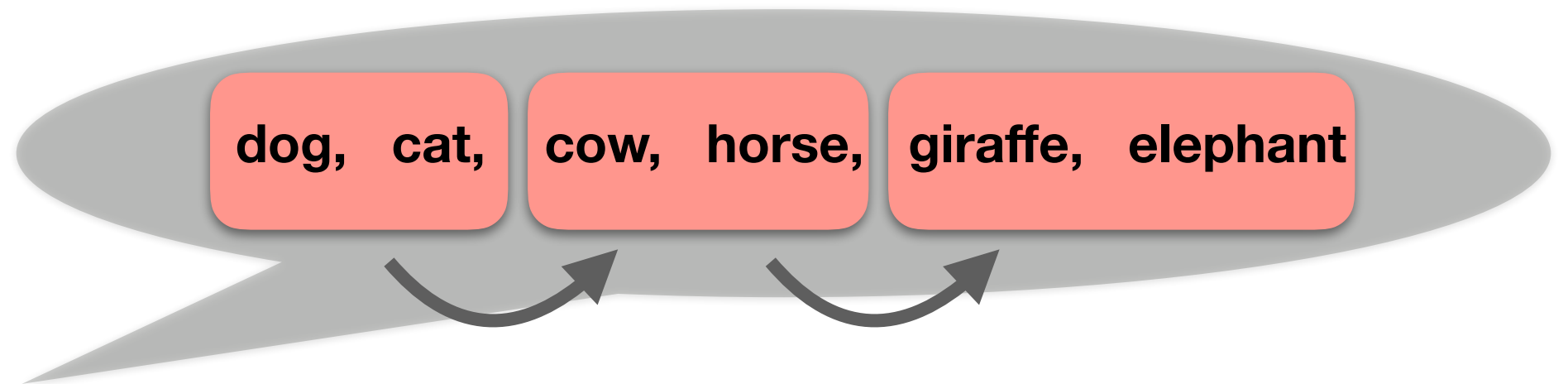
Chain



Switch



Patient



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

Control (25)

aMCI (25)

Amnesic Mild Cognitive Impairment

Memory Impairment

a+mdMCI (25)

Amnesic Multi-Domain Mild Cognitive Impairment

+ Another Cognitive Impairment

AD (25)

Alzheimer's Disease

+ Another Cognitive Impairment

+ Daily Living Activities

Materials

- **WordNet**

Knowledge-based

- **GloVe**

Word Embedding

- **Pointwise Mutual Information**

Association measure

- **Corpus**

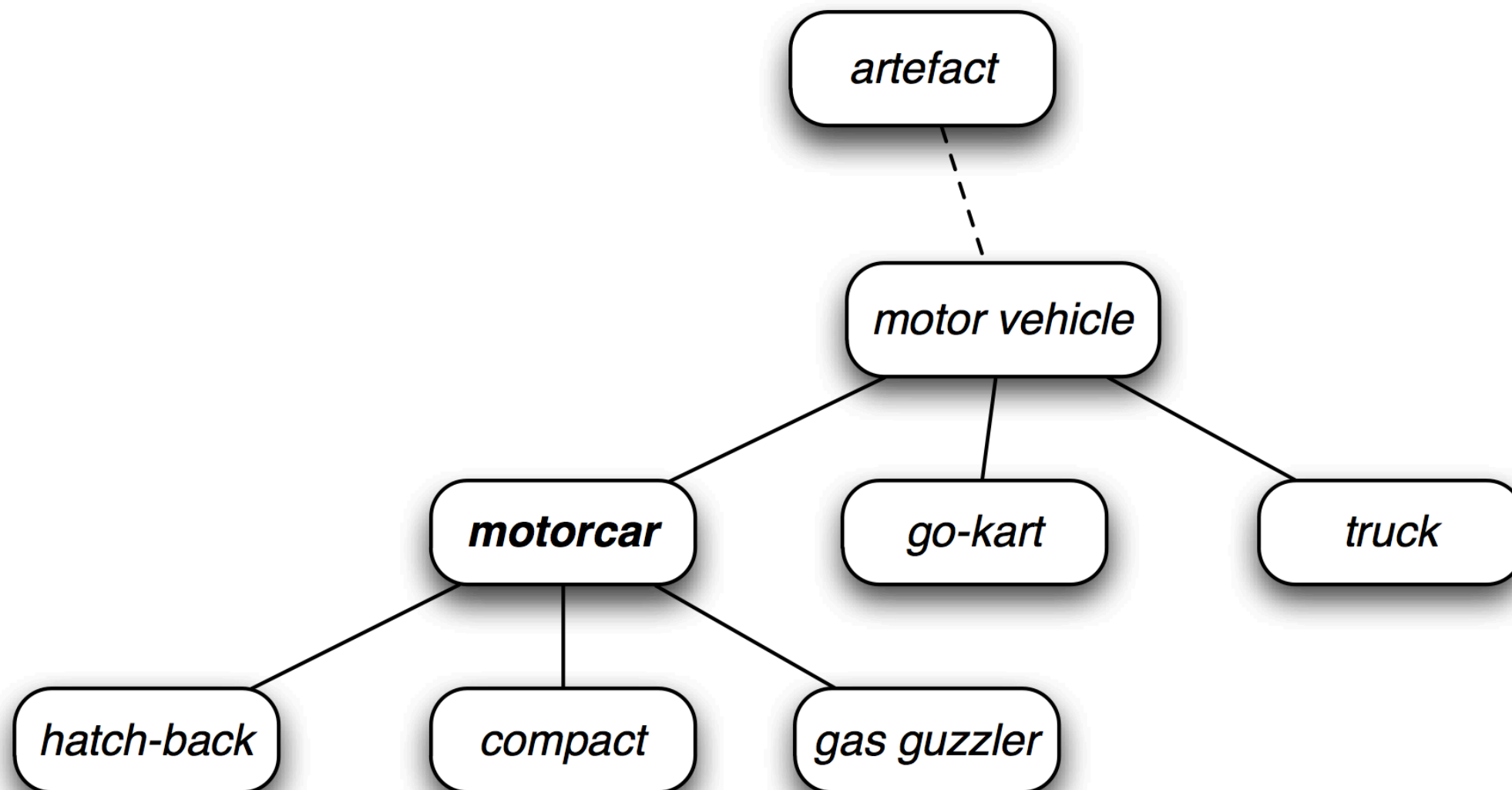
PT-BR wikipedia

44,345 types and 118,095,637 tokens

Lemmatized with PALAVRAS (Bick 2000)

Methods

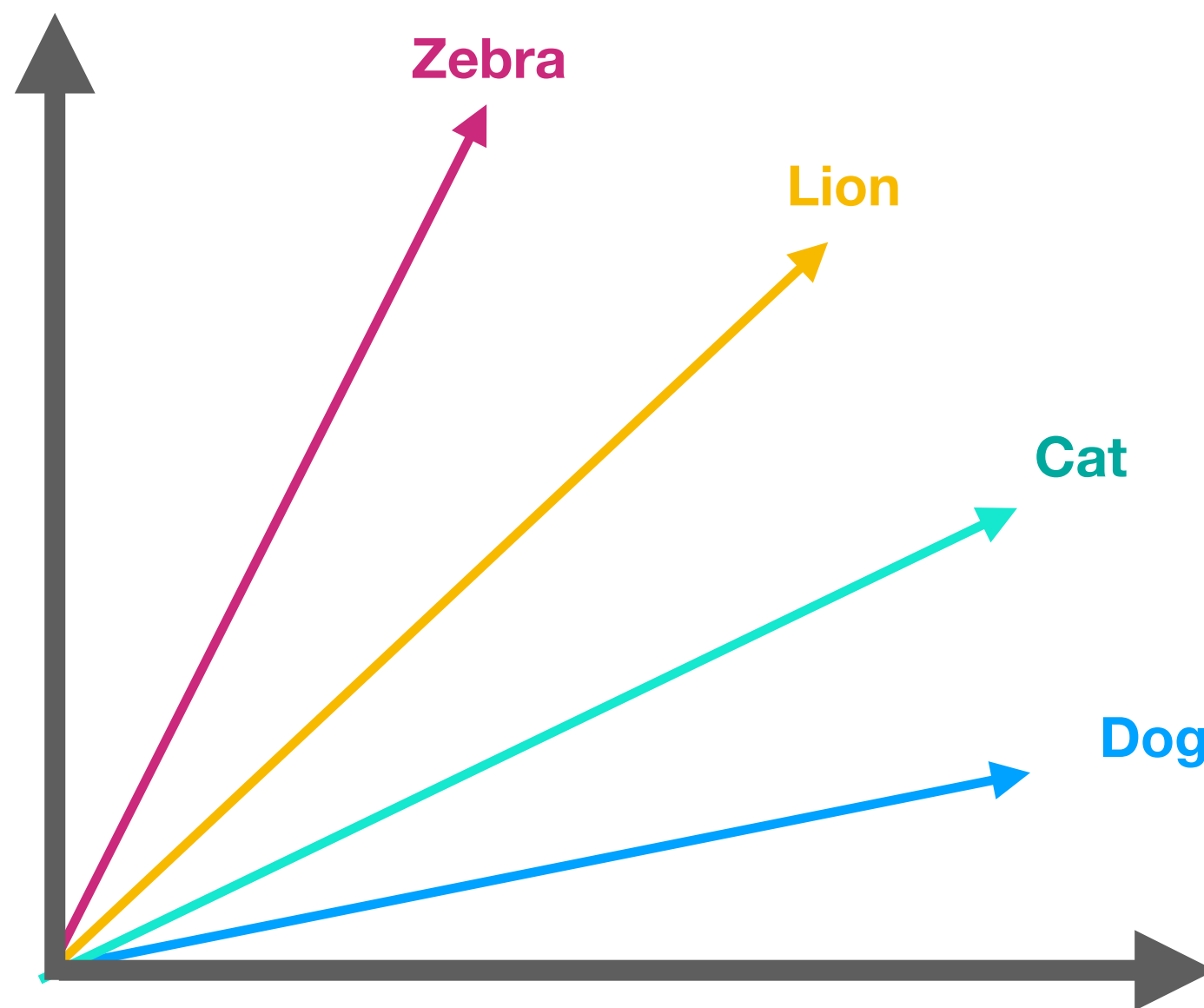
WordNet



Source: *NLTK book*, <https://www.nltk.org/book/ch02.html>

Methods

GloVe



Word

n_1
n_2
n_3
n_4
n_5
n_6
n_7
n_8
\vdots
n_n

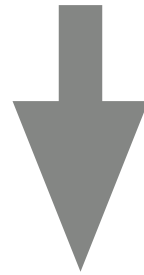
Methods

Pointwise Mutual Information

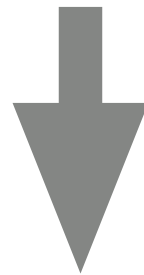
$$\text{pmi}(x; y) \equiv \log \frac{p(x, y)}{p(x)p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}.$$

Methods

Goal: find the switches in a test

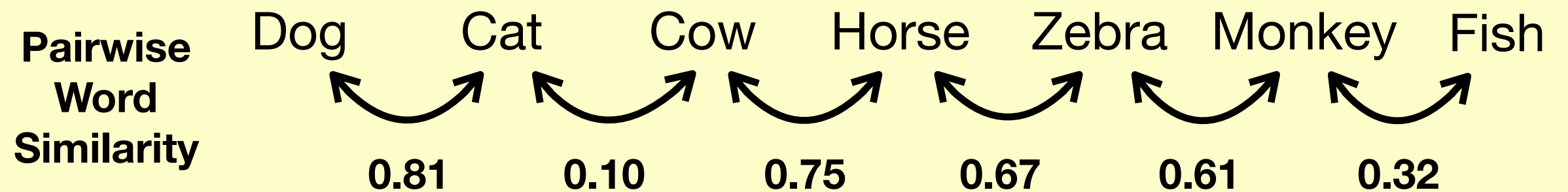


Transform the VFT in a vector



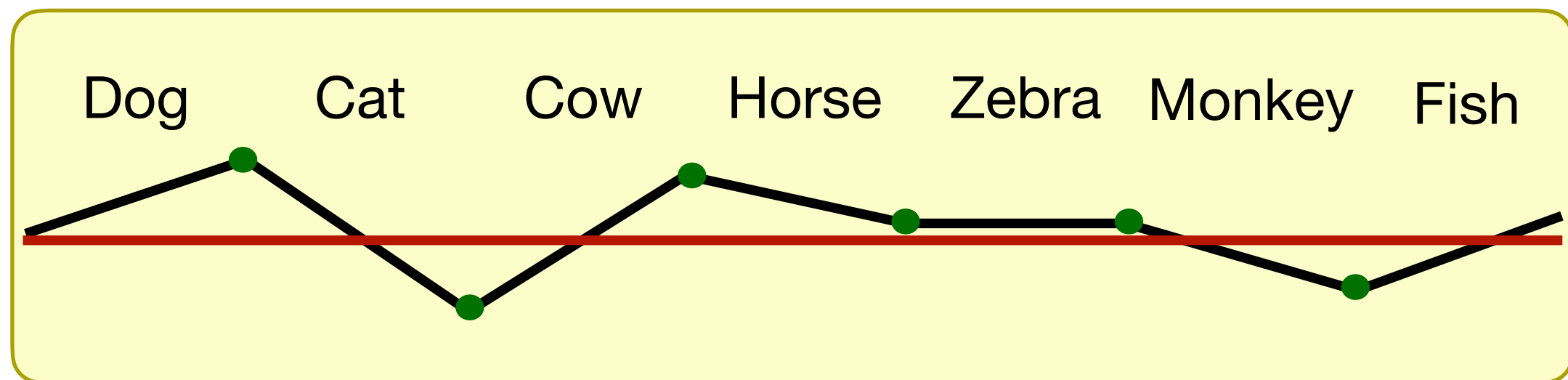
Apply a function that indicate if a pair is a switch or not

Heuristics



Heuristics

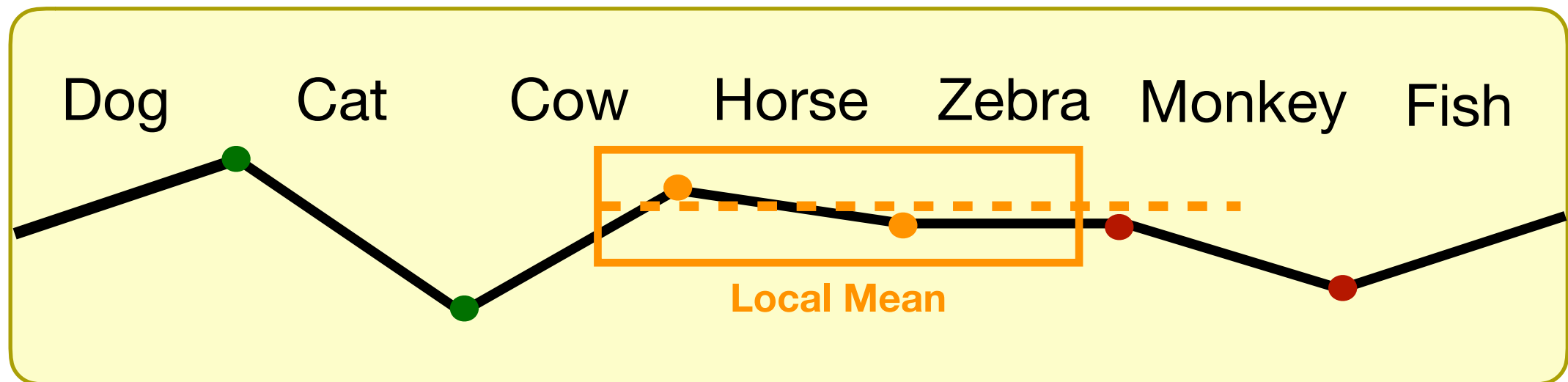
Global Mean Detection



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

Heuristics

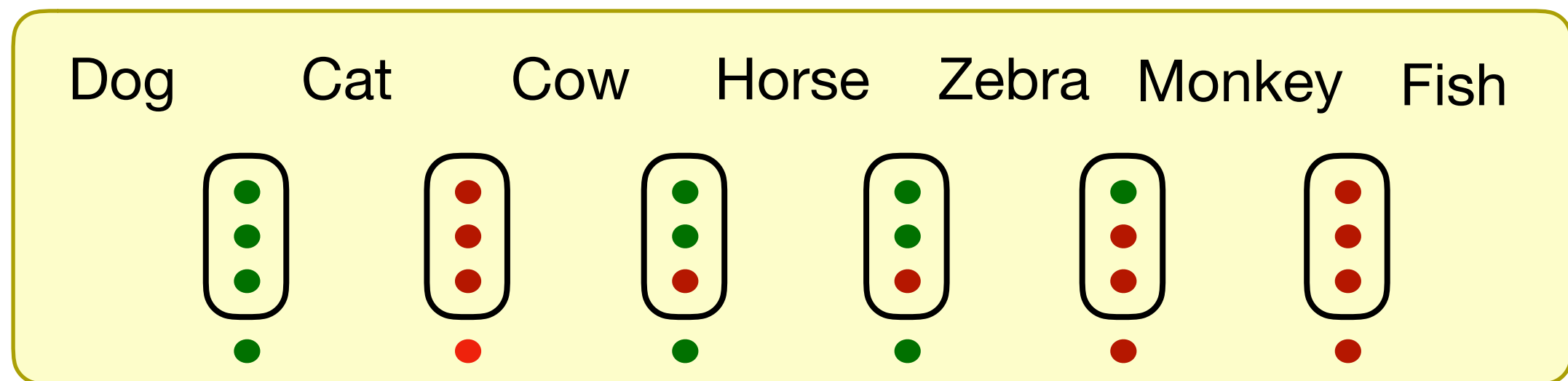
Local Mean Detection



Decision made using the mean of the last k pairs' similarity

Heuristics

Voting Detection



Decision made by voting between local and global detections

Experiments

Feature Extraction

Fraction of
Smallest Chain

Average
Chain Length

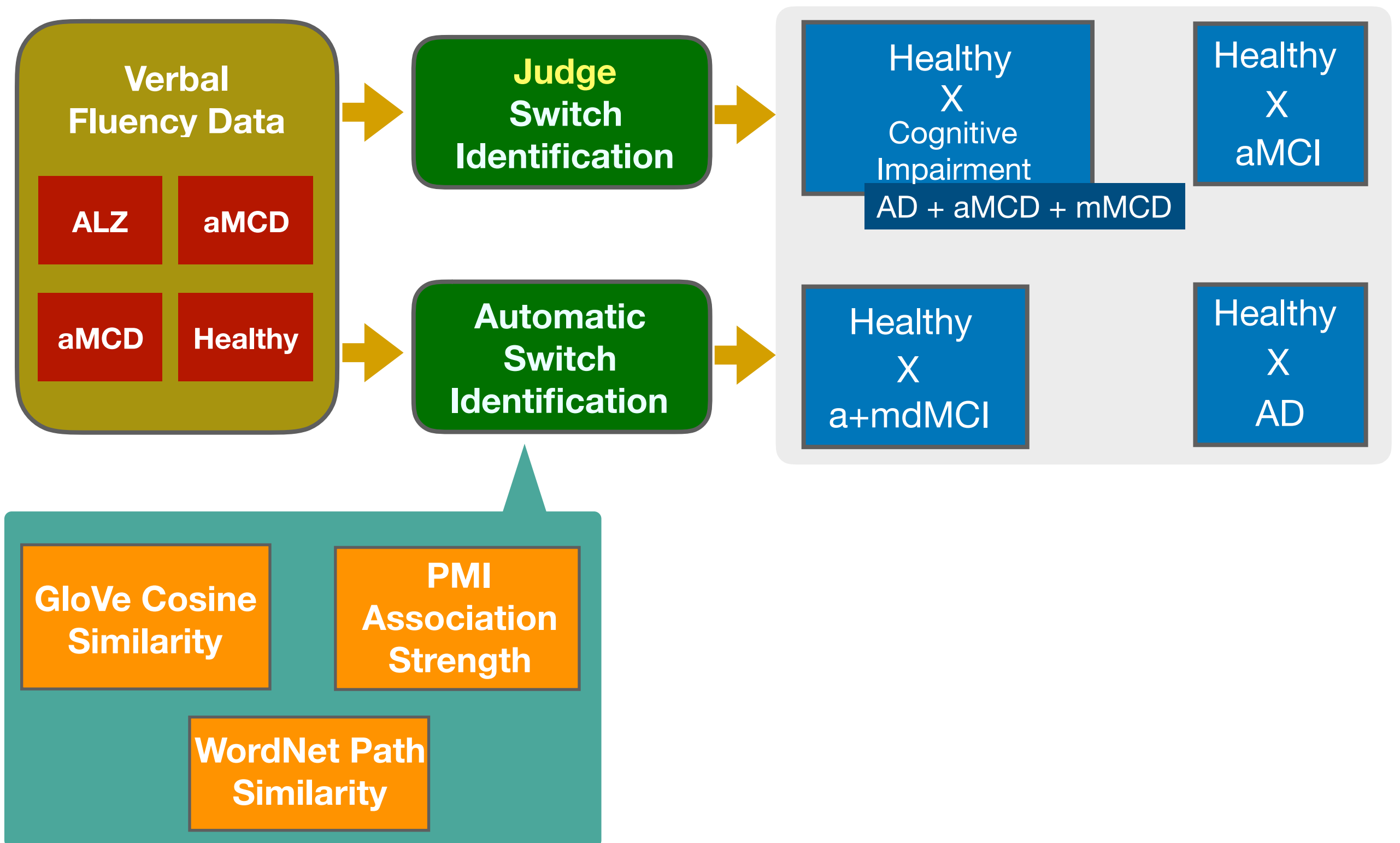
Largest
Chain Size

Number
of Switches



Classifier
Random Forest

Experiments



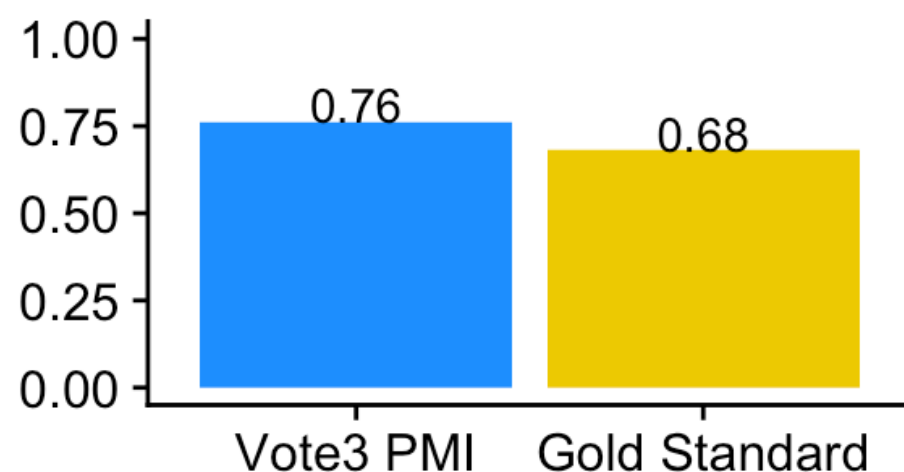
Results

Results

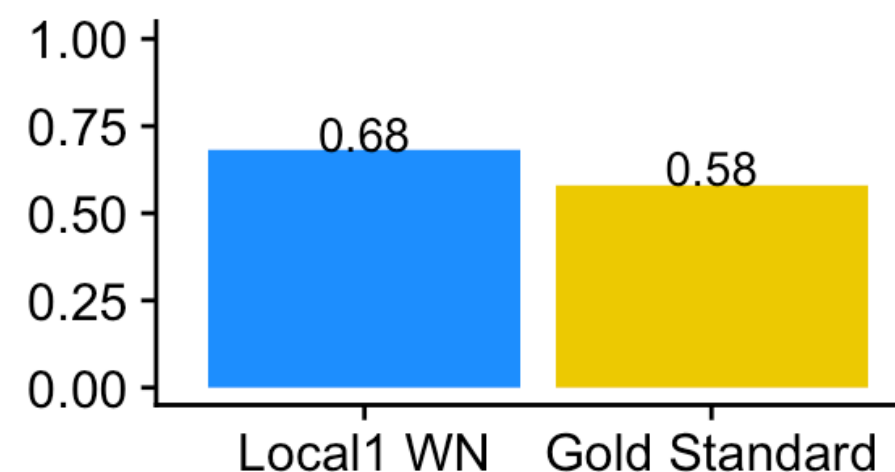
Which heuristics presented best performance?

Numeric best classifier (AUC)

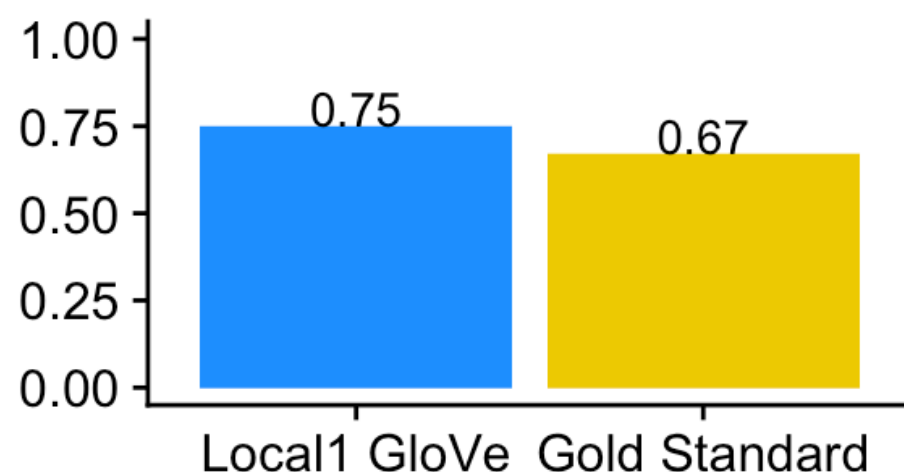
CI



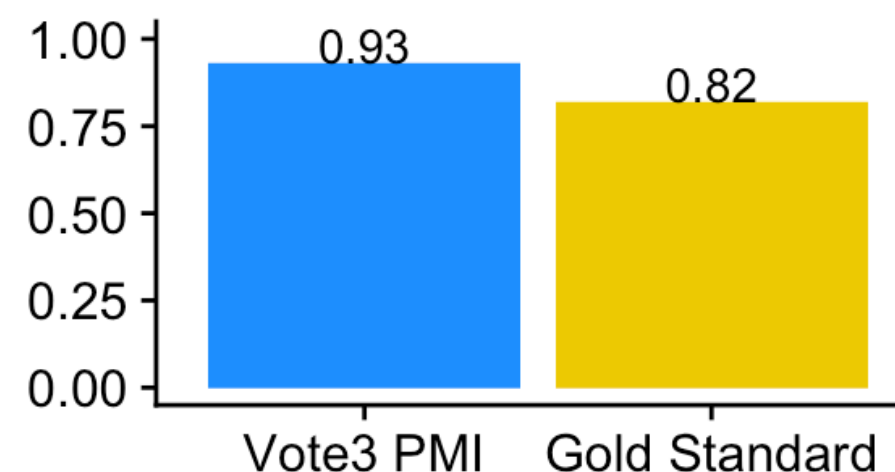
aMCI



a+mdMCI

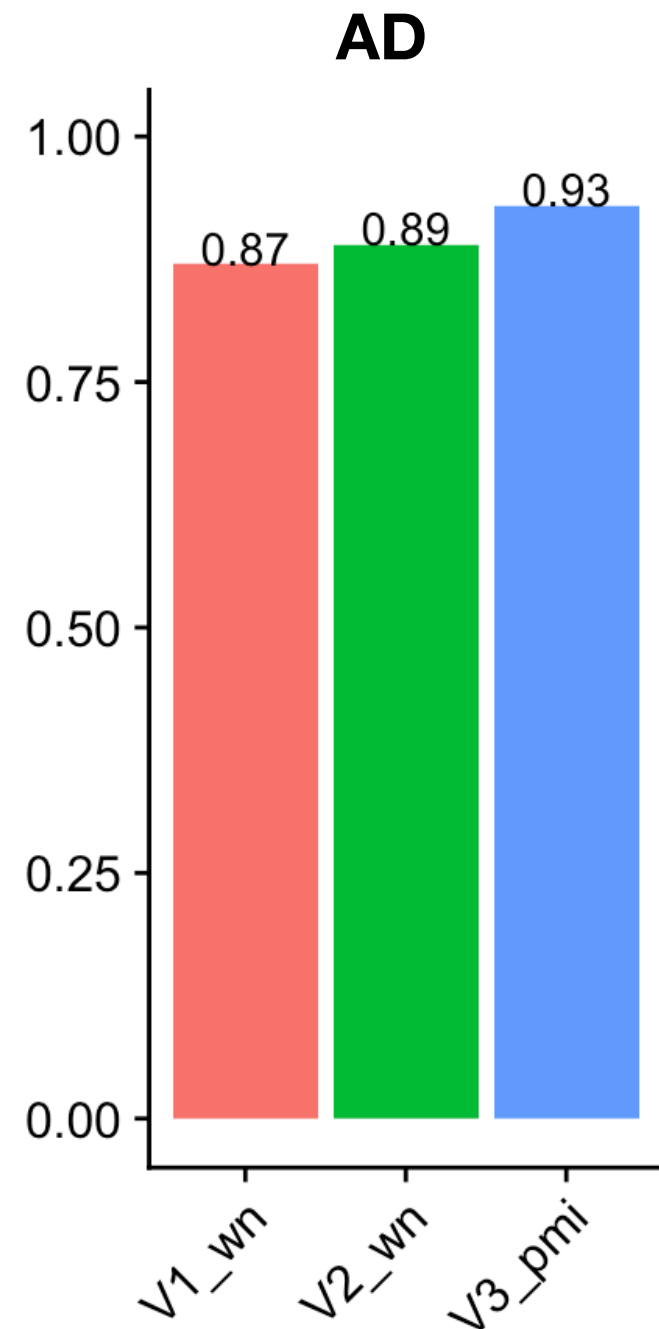
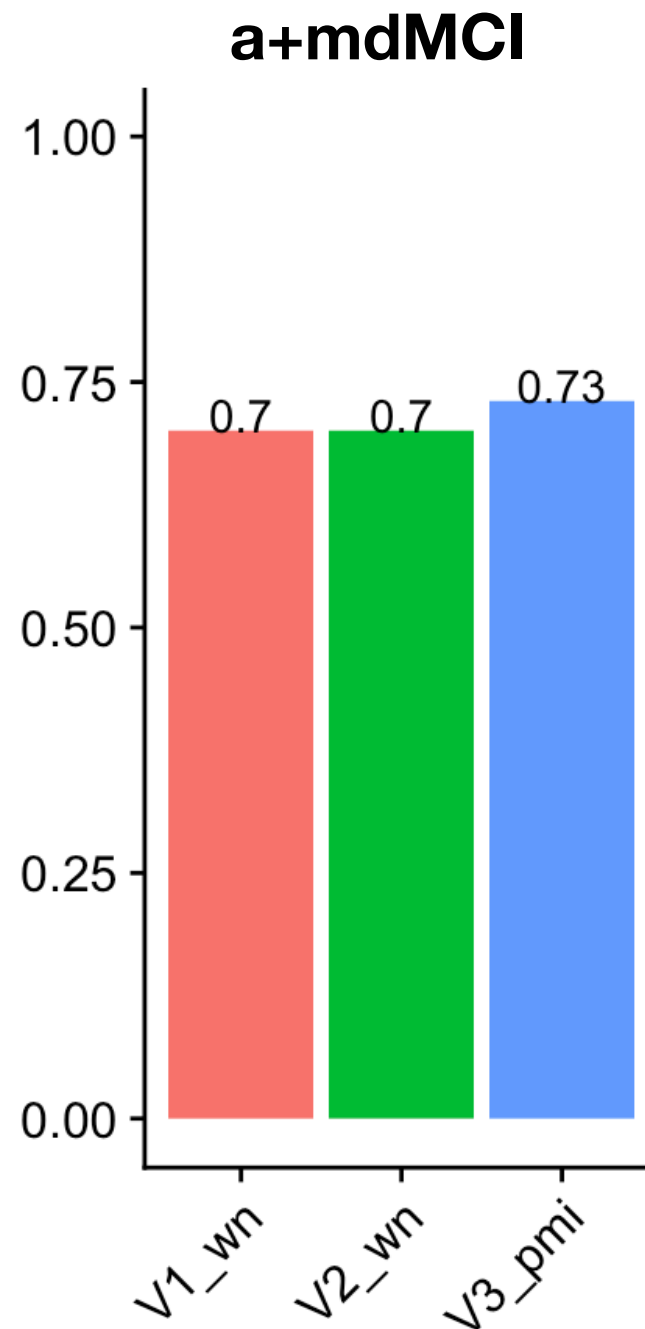
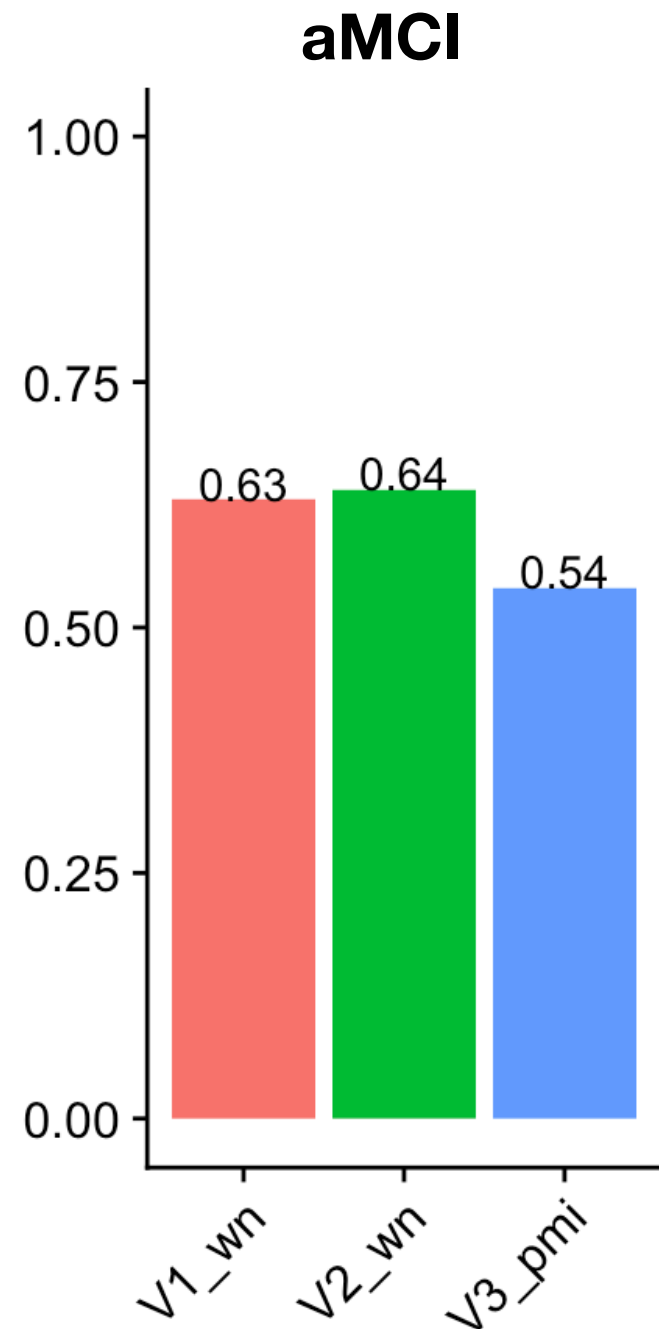


AD



Results

Which heuristics presented best performance?
Winning combinations



Conclusions

Conclusions

- 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

