Detecting neuropsychiatric conditions with semantic verbal fluency

Felipe Paula[1], Rodrigo Wilkens[2], Marco Idiart[1], and Aline Villavicencio[1,2]

[1]Federal University of Rio Grande do Sul
[2]Universite Catholique de Louvain
[2]University of Essex

1 Introduction

Semantic Verbal Fluency (SVF) [TMW+98] is a test used in several neuropsychological assessment batteries. In this test, participants are asked to produce semantically related words (e.g. animals or supermarket items) in a short period of time (e.g. one minute) avoiding repetitions. Health professionals usually look for semantic subgroups in these sequence of words, called clusters. Subjects in their recall process usually produce the clusters in an alternating way. The word that marks the alternations ares called switches.

The analysis of clusters and switches requires inspection by specialists, which are based on manually constructed taxonomies in a process that can be error prone. In this research [PWIV18], we investigate an automatic method to detect Alzheimer's disease and Mild Cognitive Impairment using SVF. Also, we assess the impact of three similarity measures in our proposed technique: WordNet path similarity, GloVe cosine similarity and PMI association strength.

2 Method

Formally the switch is a binary function $\psi(x_i)$ that operates on the sequence of N words $(w_1, w_2, \cdots, w_N)$ produced by a subject in the SVF test. There is a switch between consecutive words w_i and w_{i+1} when their similarity $x_i = s(w_i, w_{i+1})$ falls below a threshold, in which case $\psi(x_i) = 1$, otherwise $\psi(x_i) = 0$. This research explores three heuristics for the switch function.

Detection based on the global mean: The threshold is given by the average similarity of the list. $\psi_{global}(x_i) = H\left(\frac{1}{N-1}\sum_{j=1}^{N-1} x_j - x_i\right)$ where $H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ otherwise.

Detection based in the local mean: The threshold is given by the average similarity of the last k pairs of words. $\psi_k(x_i) = H\left(\frac{1}{k}\sum_{j=1}^{k} x_{i-j} - x_i\right)$.

Hibrid detection: We combine the local and global approach in a voting system where a switch is considered if it receives at least v votes from previously switch criteria. Here we consider a combination of global with locals $k = 2$ and 3: $\psi_{vot_v}(x_i) = H(\psi_{global}(x_i) + \psi_2(x_i) + \psi_3(x_i) - v)$ where v can be 1, 2 (majority voting), and 3 (total agreement).

We use Random Forest classifiers trained with the following features: the number of switches, n; the largest chain size, $c_{max} = \max(c_a)$; the average chain length, $\bar{c} = \frac{1}{n+1}\sum_{a=1}^{n+1} c_a$; the fraction of occurrence of the smallest chain, $f_{min} = \#(c_{min})/(n+1)$, where $\#(c)$ indicates the number of chains of size c in the SVF test of a participant.

To determine the effectiveness of different types of similarity measures for switch identification we examine semantic similarity from WordNet, GloVe and PMI association strength. We assess our method in SVF data from 100 eldery subjects, divided in four groups ($n = 25$): control, alzheimer (AD), amnestic mild cognitive deficit (aMCD) and multi-domain amnestic mild cognitive deficit (mMCD) [BMC+14]. The classification results are reported in terms of average area under the receiver operator characteristic curve (AUC) from 10 times 10-fold-cross validation.

3 Results and Discussion

Evaluation is carried out at two levels of granularity: a rough-grained classification for the detection of a clinical condition in general (control vs. CI group), and a fine-grained classification for one of the three conditions (aMCD, mMCD and AD groups). Table 1 displays the average AUC per heuristic for the different sources, with the highest scores shown in bold along with other scores that are not statistically different.

The last line of each subtable shows the scores obtained by training the classifiers with the gold standard manual annotation with the taxonomy used by [TMW+98] (GS in the tables).

Overall, in terms of the type of similarity both the semantic similarity (WordNet) and word association strength (PMI) were significantly better than the gold standard manual annotation for the rough-grained classification and for two of the three clinical cases (mMCD was the exception). Examining the specific groups, the lower scores for aMCD and mMCD also seem to reflect the potential progression of these condition from the control to the more severe impairments of the AD group (aMCD $<$ mMCD $<$ AD).

Among the different measures, the strict total agreement voting (ψ_{vot_3}) provides the best results with association strength for the rough-grained classification (Table 1(a)), and for the fine-grained classifications of the mMCD (Table 1(c)) and AD groups (Table 1(d)). These results suggest that a more conservative identification of switches leading to larger chains provides a better approximation for these three groups.

	(a) CI			(b) aMCD		
	WordNet	Glove	PMI	WordNet	Glove	PMI
ψ_{global}	0.64 (0.22)	0.66 (0.19)	0.66 (0.19)	0.44 (0.26)	0.56 (0.28)	**0.66 (0.28)**
ψ_1	0.65 (0.21)	0.71 (0.17)	0.68 (0.20)	**0.68 (0.25)**	0.50 (0.29)	**0.65 (0.27)**
ψ_2	0.66 (0.22)	0.66 (0.19)	0.70 (0.18)	0.50 (0.30)	0.60 (0.27)	**0.65 (0.27)**
ψ_3	**0.75 (0.19)**	0.68 (0.18)	0.66 (0.20)	0.59 (0.27)	0.58 (0.30)	0.57 (0.29)
ψ_{vot1}	**0.74 (0.17)**	0.71 (0.18)	0.62 (0.20)	**0.63 (0.27)**	**0.62 (0.27)**	0.46 (0.28)
ψ_{vot2}	**0.72 (0.19)**	0.55 (0.21)	0.69 (0.20)	**0.64 (0.28)**	0.45 (0.28)	**0.63 (0.26)**
ψ_{vot3}	**0.72 (0.18)**	0.62 (0.18)	**0.76 (0.14)**	0.61 (0.28)	0.40 (0.28)	0.54 (0.29)
GS		0.68 (0.17)			0.58 (0.27)	

	(c) mMCD			(d) AD		
	WordNet	Glove	PMI	WordNet	Glove	PMI
ψ_{global}	0.60 (0.27)	0.55 (0.27)	0.54 (0.30)	0.87 (0.17)	0.78 (0.24)	0.80 (0.23)
ψ_1	0.56 (0.30)	**0.75 (0.26)**	0.66 (0.28)	0.71 (0.25)	0.81 (0.21)	0.76 (0.22)
ψ_2	0.65 (0.28)	**0.70 (0.25)**	0.65 (0.27)	0.81 (0.21)	0.83 (0.19)	0.77 (0.25)
ψ_3	**0.71 (0.25)**	0.51 (0.27)	**0.68 (0.28)**	**0.91 (0.15)**	0.85 (0.24)	0.82 (0.20)
ψ_{vot1}	**0.70 (0.26)**	0.60 (0.30)	0.56 (0.28)	0.87 (0.22)	0.86 (0.20)	0.78 (0.23)
ψ_{vot2}	**0.70 (0.26)**	0.46 (0.26)	0.64 (0.24)	0.89 (0.16)	0.77 (0.22)	0.77 (0.21)
ψ_{vot3}	0.67 (0.24)	0.59 (0.25)	**0.73 (0.21)**	0.87 (0.18)	0.84 (0.21)	**0.93 (0.13)**
GS		0.67 (0.24)			0.82 (0.22)	

Table 1: Average scores and standard deviation for random forest classifiers trained to distinguish control from clinical groups. Switch detection with different sources of similarity (WordNet, GloVe and PMI) as well as gold standard taxonomy (GS). Control vs. Cognitive Impairment (CI), Control vs. Amnestic Mild Cognitive Deficit (aMCD), Control vs. Multi-domain Mild Cognitive Deficit (mMCD) and Control vs. Alzheimer's Disease (AD)

References

[BMC+14] Laiss Bertola, Natalia B. Mota, Mauro Copelli, Thiago Rivero, Breno Satler Diniz, Marco A. Romano-Silva, Sidarta Ribeiro, and Leandro F. Malloy-Diniz. Graph analysis of verbal fluency test discriminate between patients with alzheimer's disease, mild cognitive impairment and normal elderly controls. *Frontiers in Aging Neuroscience*, 6, jul 2014.

[PWIV18] Felipe Paula, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio. Similarity measures for the detection of clinical conditions with verbal fluency tasks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 231–235. Association for Computational Linguistics, 2018.

[TMW+98] Angela K Troyer, Morris Moscovitch, Gordon Winocur, Michael P Alexander, and Don Stuss. Clustering and switching on verbal fluency: the effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, 36(6):499–504, jun 1998.