

# Detectando condições clínicas usando semântica distribucional

Felipe S. F. Paula

Aline Villavicencio

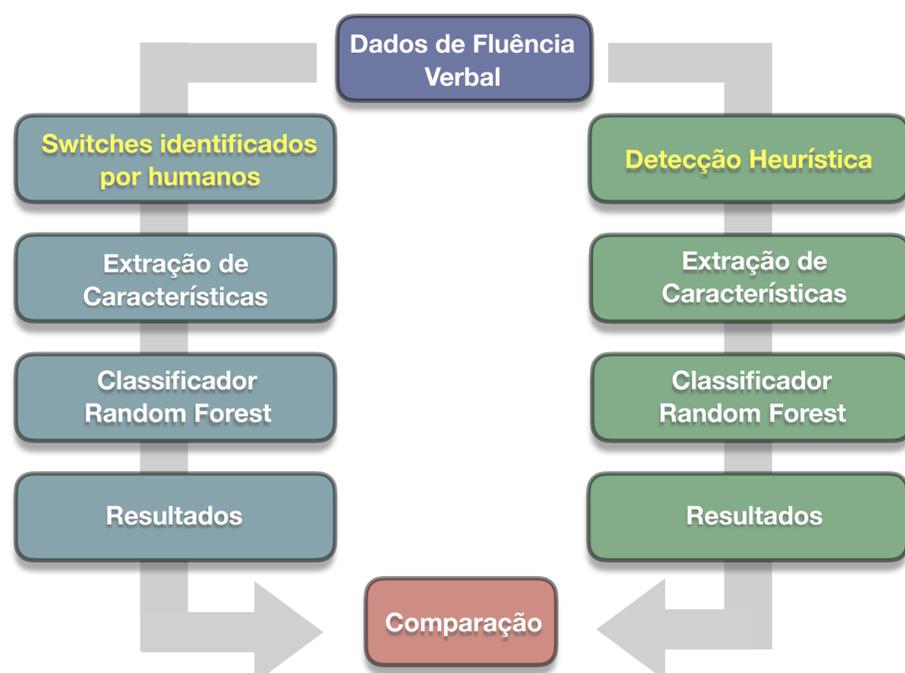
## Conceitos Utilizados

- ▶ **Semântica distribucional** é um conjunto de técnicas que usa a informação da distribuição das palavras em textos para identificar seus significados. É baseada no princípio: *you know a word by the company it keeps*. (J. R. Firth).
- ▶ Neste trabalho, algoritmos de aprendizado de máquina são usados para construir classificadores para identificação do grupo clínico com base em **testes de fluência verbal (TFV)**.
- ▶ TFV semântica são usados em avaliação neuropsicológica e ajudam a detectar diversas doenças como demências, esquizofrenia, bipolaridade, entre outras. (Zhao et. al 2014)
- ▶ Nesses testes, os indivíduos falam uma sequência de palavras pertencentes a determinada categoria.

## Resumo

- ▶ Identificamos heurísticamente subgrupos semânticos, **clusters**, e a alternância entre eles, **switches**, nos TFV.
- ▶ Além disso, obtemos clusters e switches descobertos por um humano.
- ▶ Comparamos classificadores usando características com informação humana contra os com informação heurística.

## Metodologia



## Métodos

- ▶ Treinamos modelos distribucionais no corpus da Wikipedia. Skip-gram, LexVec, GloVe e matriz de PPMI.
- ▶ Usamos um dataset de TFV com dados de **50 indivíduos idosos, divididos em dois grupos: Alzheimer (25) e Controle (25)**.
- ▶ Usando os modelos, extraímos a similaridade das palavras par a par.
- ▶ Consideramos switches todos os pares cujas similaridades estão abaixo da média dos outros pares. Chamamos isso de **detecção heurística**.



## Resultados

- ▶ Conseguimos uma **AUC 0.89** para a detecção usando GloVe.
- ▶ Esse resultado é **melhor que o classificador que utiliza detecção de switches feita pelo humano**, AUC 0.87.
- ▶ Além disso, nosso resultado é **melhor que o reportado anteriormente para esse conjunto de dados**, de AUC 0.85.

## Conclusão

- ▶ Os resultados mostram que para o caso de predição de Alzheimer, a **detecção heurística é tão boa quanto a detecção de switches por humanos**.
- ▶ Essas técnicas podem ser de **grande ajuda** para profissionais da saúde detectar doenças.