
Usando Word Embeddings para identificação de viés em notícias políticas no Brasil

Allan Sales

Depoimento de Lula para Sérgio Moro

“[...] A Folha de São Paulo, nesse mesmo período, teve 298 matérias contra o Lula e apenas 40 favoráveis. (...) O jornal O Globo, que é o mais amigo, tem 530 matérias negativas contra o Lula e 8 favoráveis. O Estadão que é mais amigo ainda, tem 318 matérias contrárias e 2 favoráveis. (...) Só o Jornal Nacional foram 18h15min nos últimos 12 meses [...]”





E para nós?

- Quão imparciais são os portais de notícias?
- Jornal x é petralha ou coxinha?
- Minha inclinação política = inclinação do jornal?

—

Como medimos a **inclinação**
política dos **portais de**
notícias brasileiros?

—

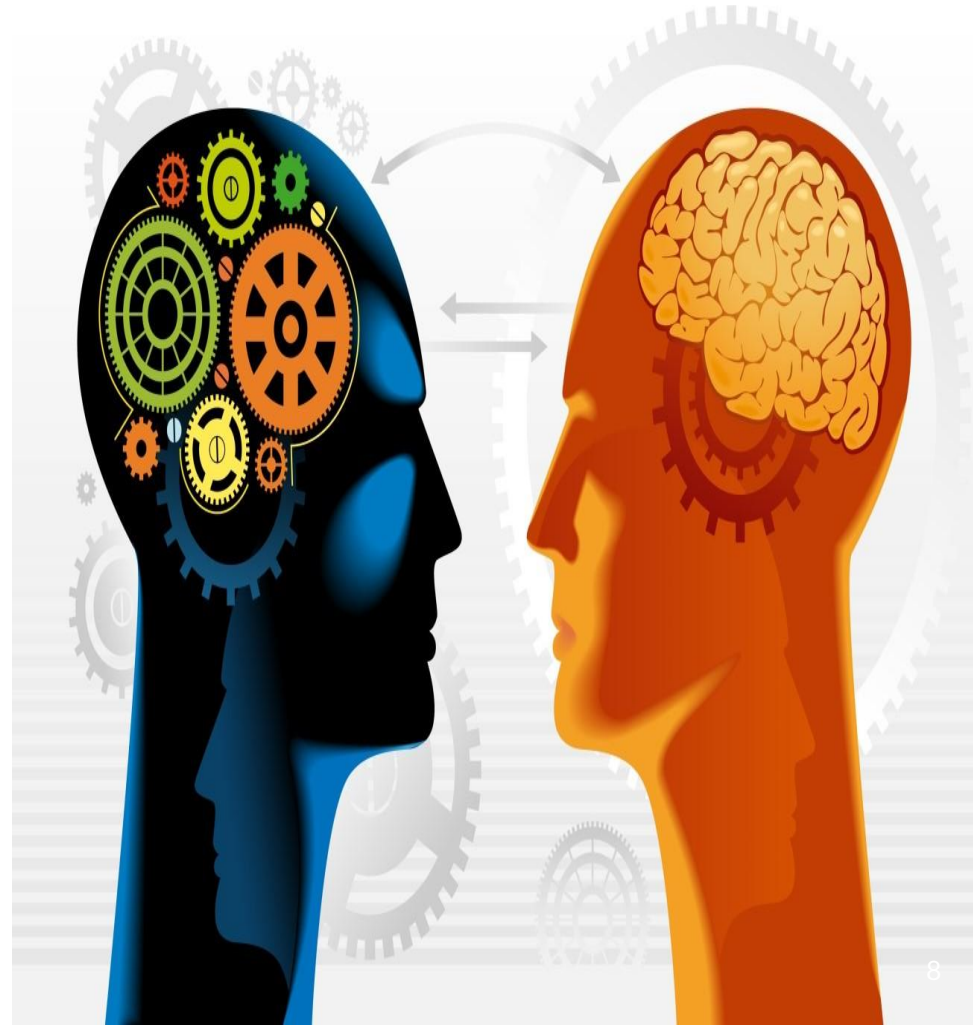
Hipótese: **textos**
apresentam os
viéses de seus
escritores.

—

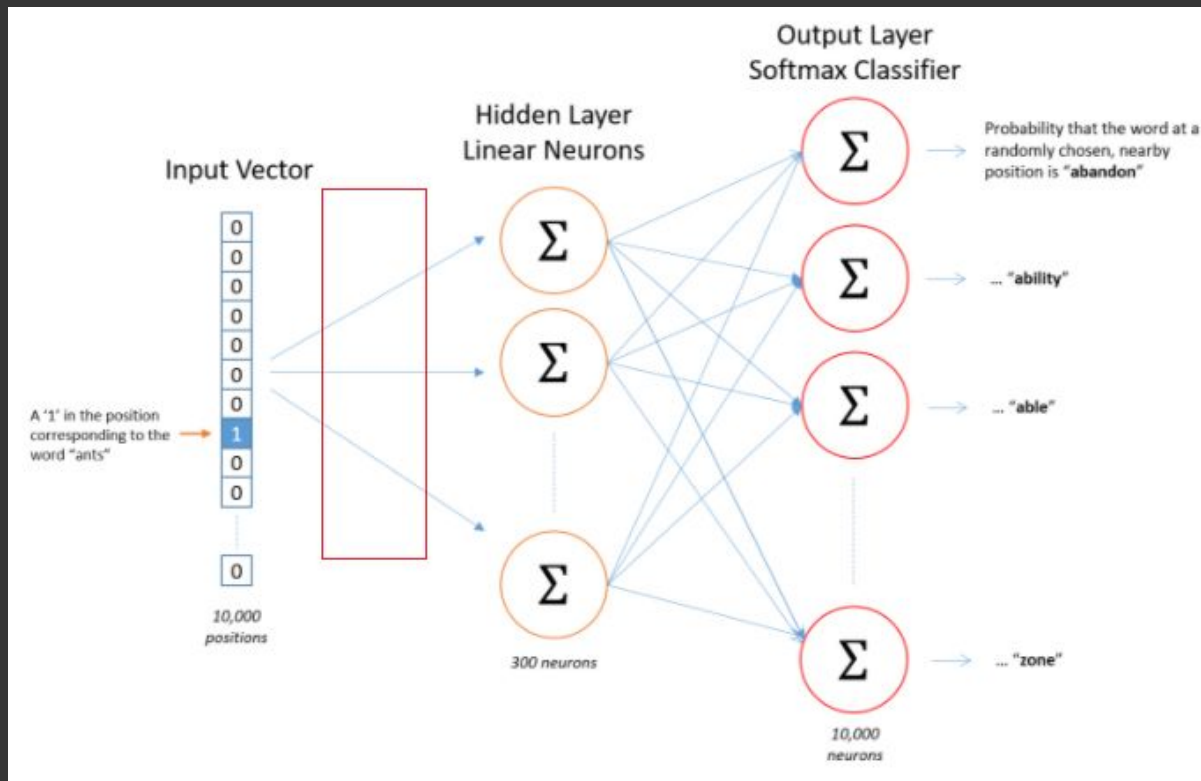
Derivar uma **métrica**
de **inclinação**
política do portal a
partir de **texto**.

Word embeddings!

O objetivo é transformar cada palavra em um vetor e fazer operações matemáticas sobre eles.



Word embedding para “ants”



Word embeddings - Eleições 2014

Jornal Estadão

- 300 camadas
- N° notícias: 2.858
- Vocabulário: 11.142
- Total de palavras: 649.335

Folha de São Paulo

- 300 camadas
- N° notícias: 3.658
- Vocabulário: 12.966
- Total de palavras: 837.218

Validação

- Dilma está para o PT assim como Aécio está para ? (PSDB)
- Marina está para PSB assim como Genro está para ? (PSOL)



Detecção de viés

(Caliskan, 2017)

Mede associação entre dois conjuntos de atributos e dois conjuntos alvos.

Alvos; Ex.:

Flores: áster, trevo, jacinto,
calêndula, papoula, azálea, açafreão,
íris, etc

Insetos: formiga, lagarta, pulga,
gafanhoto, aranha, percevejo, mosca,
etc

Atributos; Ex.:

Agradável: carícia, liberdade, saúde,
amor, paixão, amigo, prazer, etc

Desagradável: abuso, acidente,
sujeira, assassinato, doença, morte,
sofrimento, etc



Testes

- Candidatos x Temporalidade
- Candidatos x Prazer
- Gênero x Orientação partidária
(esquerda x direita)
- Candidatos x Coligação
(validação)

Existe associação implícita entre...?

Alvos	Atributos	Portal	p-valor	T. efeito
Dilma x Aécio	Duradouro x Passageiro	Estadão	0.03	1.18
		Folha SP	0.13	0.67
Dilma x Aécio	Agradável x Desagradável	Estadão	0.64	-0.25
		Folha SP	0.059	1.05
Mulheres x homens	Esquerda x Direita	Estadão	~ 0.0	1.70
		Folha SP	0.01	1.22
Dilma x Aécio	Coligação Dilma x Coligação Aécio	Estadão	~ 0.0	1.66
		Folha SP	~ 0.0	1.78

Conjuntos de palavras

Prazer:

- Agradável: carinho, liberdade, saúde, amor, paz, alegria, fiel, honesto, sorte, diploma, presente, honra, família, feliz
- Desagradável: abuso, quebra, doença, acidente, sofrimento, assalto, desastre, ódio, tragédia, cadeia, pobreza, feio, câncer, morte, prisão

Tempo:

- Passageiro: instável, variável, curto, breve, passagem
- Permantente: estável, sempre, constante, longo, persistência

Candidatos:

- Dilma: dilma, rousseff, lula, pt, temer
- Aécio: aécio, neves, fhc, psdb, aloysio

Coligação:

- Dilma: pmdb, psd, pp, pr, pdt, pros, pcdob
- Aécio: sd, pen, pmn, ptn, ptc, dem, ptb

Gênero:

- Mulheres: dilma, rousseff, marina, silva, luciana, genro
- Homens: aécio, neves, eduardo, jorge, levy, fidelix

Posicionamento:

- Esquerda: pt, psb, psd, pdt, pps, pcdob, psol
- Direita: pp, pr, prb, pros, prtb, phs, pen, ptc, dem