

# PROCESSAMENTO DE LINGUAGEM NATURAL

## EPISODE III : EMBEDDINGS INTRODUÇÃO

Alex Marino

14 de novembro de 2024

# Conteúdo

Introdução ao NLP

Word Embeddings

Modelos Clássicos

Comparação

Aplicações

Word Embeddings Contextuais

Exemplos Práticos

Desafios

Estado da Arte

Futuro

Conclusão



# Introdução ao NLP

- ▶ Definição de NLP
- ▶ Aplicações de NLP
- ▶ Desafios no NLP



# O que são Word Embeddings?

- ▶ Representação vetorial de palavras
- ▶ Capacidade de capturar o significado semântico
- ▶ Proximidade no espaço vetorial reflete similaridade semântica



# Por que usar Word Embeddings?

- ▶ Superam as representações de sacos de palavras (Bag-of-Words)
- ▶ Redução da dimensionalidade
- ▶ Captura de contextos e relações entre palavras



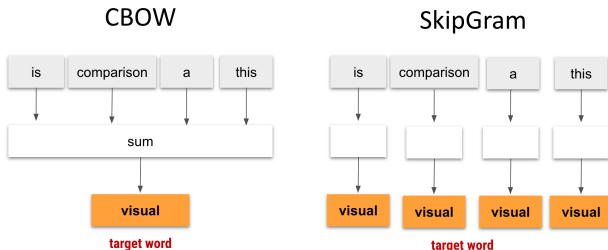
# Modelos Clássicos de Word Embeddings

- ▶ Word2Vec
- ▶ GloVe
- ▶ FastText



## Word2Vec: CBOW e Skip-gram

- ▶ CBOW (Continuous Bag of Words): Prever a palavra-alvo a partir de palavras de contexto
- ▶ Skip-gram: Prever palavras de contexto a partir de uma palavra-alvo



By: Kavita Ganesan

This is a visual comparison

Figura: Arquitetura Word2Vec

# GloVe

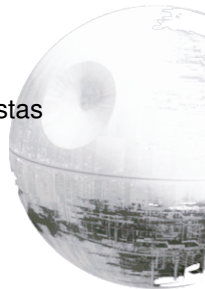
- ▶ Baseado em matrizes de coocorrência
- ▶ Captura relações globais de palavras
- ▶ Produz embeddings mais consistentes para grandes corpora





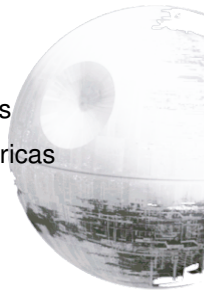
# FastText

- ▶ Modela subpalavras e morfologia
- ▶ Melhor para linguagens com palavras raras ou compostas
- ▶ Usa embeddings de n-grams



# Comparação entre Modelos

- ▶ Word2Vec: Simplicidade e eficiência
- ▶ GloVe: Embeddings mais robustos para grandes dados
- ▶ FastText: Vantagem em linguagens morfológicamente ricas



# Aplicações dos Word Embeddings

- ▶ Classificação de Sentimentos
- ▶ Tradução Automática
- ▶ Modelagem de Tópicos
- ▶ Busca Semântica



# Word Embeddings Contextuais

- ▶ Capacidade de capturar o significado das palavras em contexto
- ▶ Exemplos: ELMo, BERT



# ELMo

- ▶ Modelo baseado em redes neurais recorrentes
- ▶ Captura informações contextuais a partir de múltiplas camadas



# BERT

- ▶ Baseado na arquitetura Transformer
- ▶ Embeddings bidirecionais: Captura o contexto anterior e posterior



# Diferenças entre Embeddings Tradicionais e Contextuais

- ▶ Embeddings Tradicionais: Representações fixas para cada palavra
- ▶ Embeddings Contextuais: Representações que variam conforme o contexto



# Exemplos Práticos

- ▶ Embeddings em sistemas de recomendação
- ▶ Embeddings em sistemas de busca
- ▶ Embeddings em tradutores automáticos





# Desafios e Limitações

- ▶ Viés nos embeddings
- ▶ Necessidade de grandes quantidades de dados
- ▶ Problemas de eficiência e memória



# Estado da Arte em Word Embeddings

- ▶ Avanços com modelos pré-treinados (GPT, BERT, T5)
- ▶ Customização de embeddings para tarefas específicas



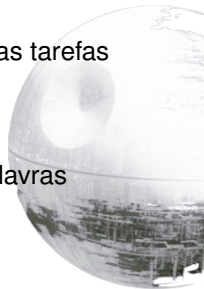
# Futuro dos Word Embeddings

- ▶ Evolução para modelos cada vez mais contextuais
- ▶ Integração com outras tecnologias de IA



# Conclusão

- ▶ Word Embeddings são fundamentais para a maioria das tarefas de NLP
- ▶ Modelos continuam evoluindo com novas arquiteturas
- ▶ Contextualização é o futuro das representações de palavras



**MAY THE**  
**SOURCE**  
**BE WITH**  
**YOU**

