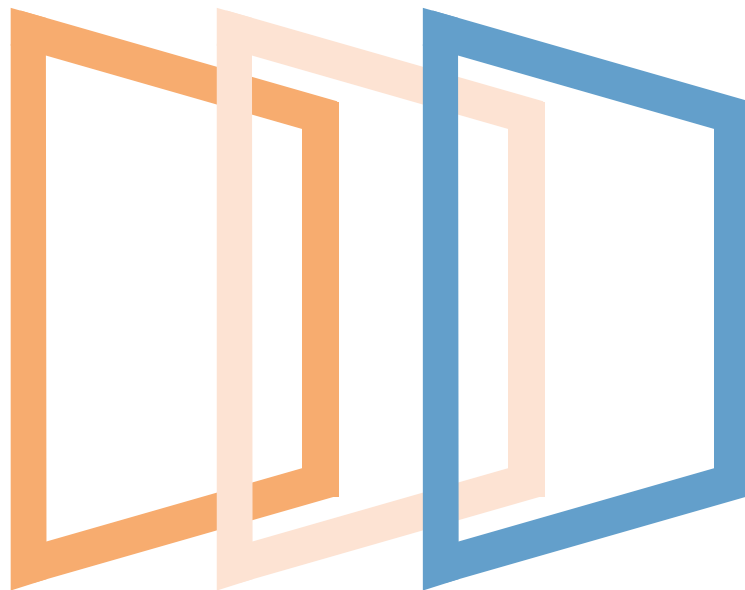


Redes Neurais Intermediário

Diego Alexandre

Práticas Tecnológicas, 02.12.2024

minsoit



An Indra company

Índice

1. Tokenization
2. Embedding
3. Alguns modelos conhecidos

Tokenization

01

Tokenization

- Transformar o texto em vetores
- Como fazer isso?

Tokenization

- Opção 1: Palavras, transformar todas as palavras do texto em vetores
 - Benefícios
 - Intuitivo
 - Problemas
 - Vocabulário grande
 - Dificuldade em lidar com palavras novas

Tokenization

- Opção 2: Transformar todos os caracteres do texto em números
 - Benefícios
 - Vocabulário pequeno
 - Sem palavras fora do vocabulário
 - Problemas
 - Perda de context dentro das palavras
 - Sequencias muito longas

Tokenization

- Opção 3: Transformar sub palavras do texto em números

Embeddings

- Transformar palavras em vetores
- De modo que palavras com significados similares fiquem próximas

Alguns modelos conhecidos

02

O que é um modelo de linguagem?

- Designa probabilidades para sequencias de palavras, encontra as palavras mais prováveis
- Pode ser
 - Generativo: Encontra as próximas palavras mais prováveis
 - Classificativo: Encontra a classificação mais provável

O que é um LLM?

- Rede Neural que processa sequencias de tamanho variavel usando um mecanismo de atenção

Como funciona um LLM?

- Tokeniza as palavras
- Cria os embeddings

Usos de um LLM

- Criação de conteúdo
- Resumir
- Responder Perguntas
- Tradução
- Classificação
- Reconhecer entidades nomeadas
- Ajuste de tom
- Geração de código

Usos de um LLM

Modelo ou família de modelos	Tamanho(# parâmetros)	Licença	Criado por	Lançamento
Falcon	7B – 40B	Apache 2.0	Technology Innovation Institute	2023
MPT	7B	Apache 2.0	MosaicML	2023
Dolly	12B	MIT	Databricks	2023
Pythia	19M – 12B	Apache 2.0	Eleuther Ai	2023
GPT-3.5	175B	Proprietário	OpenAi	2022
BLOOM	560M – 176B	RAILV1.0	BigScience	2022
FLAN-T5	80M – 540B	Apache2.0	Google	2021
BART	139M – 406M	Apache2.0	Meta	2019
BERT	109M – 335M	Apache2.0	Google	2018

Referências

- Iris Recognition with Off-the-Shelf CNN Features: A Deep Learning Perspective - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-evolution-of-the-winning-entries-on-the-ImageNet-Large-Scale-Visual-Recognition_fig1_321896881 [accessed 23 Oct, 2023]
- Application of Deep Learning in Dentistry and Implantology - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Algorithms-that-won-the-ImageNet-Large-Scale-Visual-Recognition-Challenge-ILSVRC-in_fig2_346091812 [accessed 23 Oct, 2023]
- [ImageNet classification with deep convolutional neural networks \(acm.org\)](#)
- [\[1512.03385v1\] Deep Residual Learning for Image Recognition \(arxiv.org\)](#)

- [LeNet-5 - A Classic CNN Architecture - DataScienceCentral.com](#)
- [Exploring Object Detection Applications and Benefits – DeepLobe](#)
- [\[1708.02002\] Focal Loss for Dense Object Detection \(arxiv.org\)](#)
- [Mean Average Precision \(mAP\) Using the COCO Evaluator – PyImageSearch](#)
- [http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf](#)
- [https://kharshit.github.io/blog/2019/02/15/autoencoder-downsampling-and-upsampling](#)
- [\[1505.04597\] U-Net: Convolutional Networks for Biomedical Image Segmentation \(arxiv.org\)](#)

Referências

- [\[1612.03144\] Feature Pyramid Networks for Object Detection \(arxiv.org\)](#)
- [\[1706.05587v3\] Rethinking Atrous Convolution for Semantic Image Segmentation \(arxiv.org\)](#)
- [\[1606.00915\] DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs \(arxiv.org\)](#)
- [\[1802.02611\] Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation \(arxiv.org\)](#)

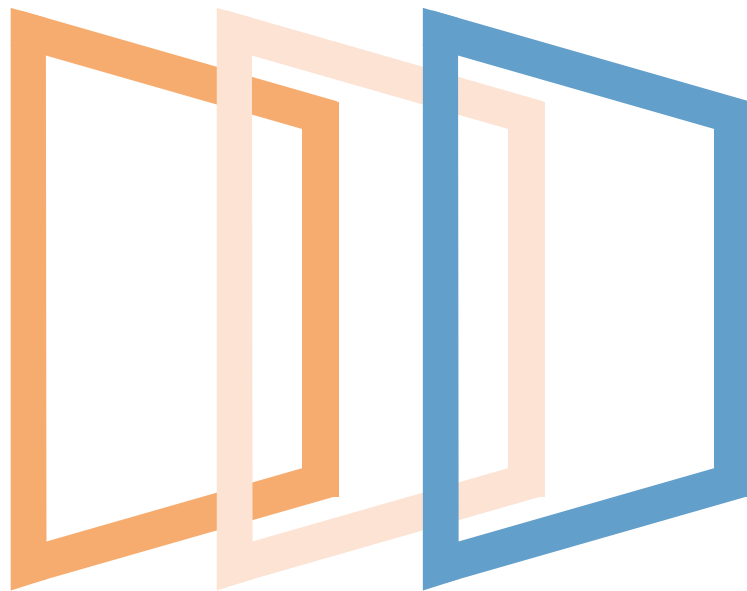
Redes Neurais Convolucionais

Thaís Ratis

Diego Alexandre

Práticas Tecnológicas, 26.10.2023

minsoit



An Indra company