

IA Generativa

Glossário do Curso: Modelos de IA para PLN

Bem-vindo! Este glossário em ordem alfabética contém muitos dos termos deste curso. Este glossário abrangente também inclui termos adicionais reconhecidos pela indústria que não são usados nos vídeos do curso. Esses termos são essenciais para você reconhecer ao trabalhar na indústria, participar de grupos de usuários e em outros programas de certificação.

Tempo estimado de leitura: 4 minutos

Termo	Definição
Bag-of-words	Uma representação que retrata um documento como o agregado ou a média de vetores one-hot codificados. Representa documentos como um conjunto de palavras e considera a frequência de ocorrência de uma palavra dentro do documento.
Modelo bi-grama	Um modelo de probabilidade condicional com tamanho de contexto um, o que significa que você considera apenas as palavras adjacentes na sequência para prever a próxima.
Vetor de contexto	Produto do tamanho do contexto e do tamanho do vocabulário. Normalmente, esse vetor não é computado diretamente, mas é construído concatenando os vetores de incorporação.
Bag of words contínua (CBOW)	Um modelo que utiliza palavras de contexto para prever uma palavra-alvo e gerar sua incorporação.
Perda de entropia cruzada	Uma métrica usada para medir o desempenho de um modelo de classificação. A saída é um número entre 0 e 1. Quanto menor o número, melhor o modelo.
Carregador de dados	Componente de aplicação que permite o agrupamento e embaralhamento eficientes de dados, o que é essencial para treinar redes neurais. Permite o pré-processamento em tempo real, otimizando o uso da memória. Carregadores de dados são importantes para gerenciar grandes conjuntos de dados de forma eficiente durante o treinamento do modelo.
Conjunto de dados	Uma coleção de amostras de dados e seus rótulos.
Camada de incorporação	Uma camada que aceita índices de tokens e produz vetores de incorporação.
Ajuste fino	Ajustando um modelo pré-treinado para melhorar o desempenho em uma tarefa ou conjunto de dados específico. Isso faz com que o modelo gere conteúdo mais preciso e contextualmente relevante.

Termo	Definição
Unidades recorrentes com portão (ou GRUs)	Melhorias populares em redes neurais recorrentes (RNN) com um mecanismo de portão para controlar o fluxo de informações dentro da rede. Elas são semelhantes às memórias de longo prazo (LSTM), mas podem ser treinadas rapidamente.
Hiperparâmetros	Configurações de um modelo de rede neural que são externas ao modelo e definem aspectos como o comportamento durante o treinamento.
Grandes modelos de linguagem (LLMs)	Modelos fundamentais que utilizam IA e aprendizado profundo com vastos conjuntos de dados para gerar texto, traduzir idiomas e criar vários tipos de conteúdo. Eles são chamados de grandes modelos de linguagem devido ao tamanho do conjunto de dados de treinamento e ao número de parâmetros.
Parâmetros aprendíveis	Os pesos e vieses em uma rede neural que são otimizados durante o treinamento de um modelo.
Taxa de aprendizado	Um hiperparâmetro que determina quão rapidamente ou lentamente a rede neural aprende com os dados. Regula o tamanho do passo no processo de otimização.
Logits	Saídas brutas e não normalizadas de uma rede neural antes que a função de ativação seja aplicada.
Memória de longo prazo (ou LSTMs)	Melhorias populares em redes neurais recorrentes (RNN) eficazes para tarefas que envolvem dados de séries temporais extensas, como processamento de linguagem natural (PLN).
Função de perda	Uma medida que representa a diferença entre os valores previstos por um modelo e os valores reais nos dados de treinamento.
Amostragem de Monte Carlo	Uma técnica estatística que envolve a geração de amostras aleatórias a partir de uma distribuição de probabilidade. É especialmente benéfica ao lidar com sistemas que envolvem incerteza.
Processamento de linguagem natural (PLN)	O subcampo da inteligência artificial (IA) que lida com a interação de computadores e humanos na linguagem humana. Envolve a criação de algoritmos e modelos que ajudarão os computadores a entender e compreender a linguagem humana e gerar texto contextualmente relevante na linguagem humana.
Redes neurais	Modelos computacionais inspirados na estrutura do cérebro humano. Um modelo de rede neural consiste em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída.
Modelo n-grama	Modelo de linguagem que analisa sequências de ‘n’ itens consecutivos, muitas vezes palavras, para prever padrões ou frases que ocorrem em um texto. O modelo n-grama permite um tamanho de contexto arbitrário.
NLTK	Uma biblioteca Python usada em processamento de linguagem natural (PLN) para tarefas, como tokenização e processamento de texto.

Termo	Definição
Codificação one-hot	O método usado para converter dados categóricos em vetores de características que uma rede neural pode entender.
Perplexidade	Métrica para avaliar a eficiência de grandes modelos de linguagem (LLMs) e modelos de IA generativa. Na modelagem de linguagem, a perplexidade pode ser vista como uma medida de quão surpreso ou incerto o modelo está ao prever a próxima palavra em uma sequência. Valores de perplexidade mais baixos indicam melhor desempenho dos modelos de linguagem.
PyTorch	Um framework de aprendizado profundo dinâmico desenvolvido pelo laboratório de Pesquisa em IA do Facebook. É uma biblioteca baseada em Python bem conhecida por sua facilidade de uso, flexibilidade e gráficos de computação dinâmicos.
Redes neurais recorrentes (ou RNNs)	Redes neurais artificiais que usam dados sequenciais ou de séries temporais. Você pode usar RNNs para resolver problemas relacionados a dados com uma ordem natural ou dependências baseadas no tempo. Elas têm loops em sua arquitetura, permitindo que a informação persista ao longo do tempo, tornando-as adequadas para o processamento de dados sequenciais.
Modelo sequência para sequência	Arquitetura de rede neural, onde tanto a entrada quanto a saída são sequências de dados. É usado em tradução automática, como converter frases em inglês para francês.
Modelo skip-gram	Um modelo de incorporação de palavras que prevê palavras de contexto circundantes a partir de uma palavra-alvo específica. Um modelo skip-gram é o reverso do modelo bag of words contínua (CBOW).
Incorporação de palavras	Representação de palavras como vetores densos, capturando sua relação com base no contexto.
Word2vec	O grupo de modelos que produzem incorporações ou vetores de palavras, que são representações numéricas capturando a essência das palavras. É a forma abreviada de “word to vector.”

Registro de Alterações

Data	Versão	Alterado por	Descrição da Alteração
15-01-2024	0.1	Gagandeep	Versão inicial criada

—>