

Glossário: Avanço do Ajuste Fino de IA Generativa para LLMs

Bem-vindo! Este glossário em ordem alfabética contém muitos termos usados neste curso. Compreender esses termos é essencial ao trabalhar na indústria, participar de grupos de usuários e participar de outros programas de certificação.

Termo	Definição
Argmax/Argmin	O argumento do máximo/mínimo de uma função, ou seja, o valor da variável de entrada que produz o maior/menor valor de saída.
AutoModelForCausalLM	Uma classe da biblioteca Hugging Face Transformers usada para carregar modelos de linguagem causal pré-treinados, como o GPT-2, para tarefas de geração de texto.
AutoModelForCausalLMWithValueHead	Uma extensão da classe AutoModelForCausalLM na Hugging Face para aprendizado por reforço, incluindo uma cabeça de valor usada para estimar a função de valor, crucial para modelos como PPO.
Beam search	Um algoritmo de busca que expande as sequências de tokens mais promissoras em cada passo na geração de sequências, usado para melhorar a qualidade das saídas em modelos de linguagem ao considerar várias possibilidades simultaneamente.
Parâmetro Beta (β)	Um hiperparâmetro em aprendizado por reforço que controla o equilíbrio entre a política atual e o modelo de referência, impactando a troca entre exploração e exploração na otimização da política. No contexto do DPO, atua como o parâmetro de temperatura para a perda do DPO.
Modelo Bradley-Terry	Um modelo probabilístico usado para classificar e comparar diferentes itens ou escolhas, frequentemente utilizado para modelar preferências pareadas, onde a probabilidade de um item ser preferido em relação a outro é baseada em suas respectivas pontuações.
Solução em forma fechada	Uma expressão analítica explícita para a solução de um problema que não requer métodos iterativos ou numéricos para ser resolvida.
Função coladora	Uma função que organiza e agrupa dados de entrada em um formato adequado para processamento por modelos de aprendizado de máquina, especialmente em cenários de aprendizado por reforço.
Função de custo	Uma função que representa o custo associado a um conjunto específico de parâmetros em um problema de otimização. É usada para guiar os modelos em direção a um melhor desempenho, minimizando o custo durante o treinamento.
Coleta de dados	O processo de reunir e preparar conjuntos de dados, particularmente conjuntos de dados de preferência, para uso no treinamento de modelos como aqueles que utilizam otimização de preferência direta (DPO).
Conjunto de dados	Uma coleção de dados usada para treinar, validar e testar modelos de aprendizado de máquina. Neste contexto, refere-se especificamente ao conjunto de dados IMDB usado para análise de sentimentos.
Tokenização de conjunto de dados	O processo de converter dados de texto bruto em IDs de tokens que podem ser processados por modelos de aprendizado de máquina, particularmente modelos de linguagem.
Otimização de preferência direta (DPO)	Uma técnica de otimização que aproveita comparações pareadas (preferências) em vez de recompensas ou pontuações explícitas para treinar modelos, especialmente útil em cenários onde atribuir pontuações numéricas precisas é desafiador.

Termo	Definição
Distribuição (em ML)	Uma função que mostra todos os valores possíveis de um conjunto de dados e com que frequência eles ocorrem. No contexto de modelos de linguagem, refere-se à distribuição de probabilidade de diferentes respostas possíveis dadas uma consulta de entrada.
Ajuste fino	O processo de adaptar um modelo pré-treinado a uma tarefa ou conjunto de dados específico, continuando o processo de treinamento em novos dados. O ajuste fino permite que os modelos alcancem melhor desempenho na tarefa específica aproveitando o conhecimento existente da pré-formação.
Hugging Face	Uma plataforma que fornece ferramentas, bibliotecas e recursos para construir, treinar e implantar modelos de aprendizado de máquina, especialmente aqueles baseados em transformers como o GPT-2.
Conjunto de dados IMDB	Um conjunto de dados contendo 50.000 críticas de filmes usadas para análise de sentimentos, comumente utilizado para treinar modelos para classificar críticas como positivas ou negativas.
Inferência	O processo de usar um modelo de aprendizado de máquina treinado para fazer previsões ou gerar saídas com base em novos dados de entrada. No contexto de modelos de linguagem, a inferência refere-se à geração de texto ou à realização de previsões usando o modelo treinado.
Divergência de Kullback-Leibler (KL)	Uma medida de como uma distribuição de probabilidade diverge de uma segunda distribuição de probabilidade de referência. É frequentemente usada para garantir que a nova política permaneça próxima da antiga durante o treinamento em aprendizado por reforço.
Modelo de linguagem	Um modelo que prevê a probabilidade de uma sequência de palavras. É usado em várias aplicações, incluindo a geração de respostas em IA conversacional com base em uma consulta de entrada.
LengthSampler	Um método usado para variar os comprimentos de texto para processamento de dados em modelos de aprendizado de máquina, melhorando a robustez e simulando condições de treinamento realistas ao gerenciar os comprimentos de texto de entrada.
Truque da log-derivada	Uma técnica matemática usada para calcular o gradiente de uma função quando a função é dada em uma forma de expectativa, frequentemente utilizada em aprendizado por reforço para otimizar políticas.
Adaptação de baixo rank (LoRA)	Uma técnica para ajuste fino eficiente em parâmetros, particularmente em modelos de transformers, que adiciona matrizes de baixo rank treináveis a cada camada de um modelo pré-treinado para reduzir o custo computacional do treinamento e tornar o ajuste fino mais eficiente em termos de memória.
Função de perda	Uma função que mede a diferença entre os resultados previstos de um modelo e os valores alvo reais. É usada para guiar a otimização do modelo minimizando essa diferença durante o treinamento.
Tokens máximos e mínimos	Parâmetros que definem o número máximo ou mínimo de tokens gerados em uma sequência, usados para controlar o comprimento das saídas em modelos de linguagem.
Função objetivo	Uma função matemática que representa o objetivo de um problema de otimização, frequentemente usada em aprendizado de máquina para guiar os modelos em direção a um melhor desempenho minimizando ou maximizando essa função.
Função Omega (ω)	Uma notação usada para descrever a distribuição detalhada da política em aprendizado por reforço, representando como as probabilidades de

Termo	Definição
	uma resposta são calculadas com base nos tokens anteriores na sequência de entrada.
Otimização	O processo de ajustar os parâmetros de um modelo para melhorar seu desempenho, tipicamente minimizando uma função de perda ou maximizando uma recompensa. No DPO, a otimização envolve maximizar a log-verossimilhança da perda do DPO.
Função de partição	Uma função matemática que soma todos os estados ou resultados possíveis de um sistema, frequentemente usada em mecânica estatística e na normalização de distribuições de probabilidade em aprendizado de máquina, especialmente em aprendizado por reforço, onde o número de resultados possíveis pode crescer exponencialmente.
Política π	Em aprendizado por reforço, a política que define a distribuição de probabilidade sobre ações dadas um estado, frequentemente denotada como π . É o modelo que é otimizado para alcançar o melhor desempenho em tarefas de tomada de decisão.
Lista de saídas do pipe	Uma lista que armazena as saídas de um pipeline na Hugging Face, particularmente no contexto de análise de sentimentos, onde contém as pontuações de sentimento para as respostas geradas.
Gradiente de política	Um método em aprendizado por reforço que otimiza a política diretamente maximizando a recompensa esperada.
Classe de configuração PPO	Uma classe usada para especificar o modelo e a taxa de aprendizado para o treinamento de otimização de política proximal (PPO), definindo configurações essenciais para o treinamento de modelos.
Treinador PPO	Um treinador especializado em aprendizado por reforço que processa amostras de consulta, otimiza políticas de chatbot e lida com tarefas complexas para garantir respostas de alta qualidade.
Otimização de política proximal (PPO)	Um algoritmo de aprendizado por reforço que otimiza a política de um agente garantindo que as atualizações não sejam muito drásticas, estabilizando assim o processo de treinamento.
Aprendizado por reforço a partir de feedback humano (RLHF)	Uma técnica de aprendizado por reforço onde o feedback humano é usado para guiar o processo de aprendizado dos modelos, particularmente útil na otimização de grandes modelos de linguagem para tarefas como chatbots e sistemas de recomendação.
Penalidade de repetição	Um parâmetro que penaliza sequências repetidas de tokens durante a geração de texto, incentivando saídas mais diversas e reduzindo a probabilidade de gerar conteúdo repetitivo.
Modelo de referência	Um modelo pré-treinado usado como ponto de referência ou comparação em treinamento ou otimização adicional, particularmente em tarefas de aprendizado por reforço.
Função de recompensa	Em aprendizado por reforço, uma função que fornece feedback sobre a qualidade das ações tomadas por um modelo, orientando o processo de aprendizado ao indicar quais ações levam a recompensas mais altas.
Distribuição ponderada por recompensa	Uma distribuição de probabilidade que foi ajustada com base nas recompensas obtidas, usada em aprendizado por reforço para guiar a otimização de políticas em direção a ações que geram recompensas mais altas.
Rollout	O processo pelo qual um modelo gera diferentes respostas para uma determinada consulta, usado em aprendizado por reforço para avaliar a eficácia das políticas. Em bibliotecas como a Hugging Face, o termo

Termo	Definição
	pode diferir ligeiramente, mas geralmente se refere às múltiplas saídas possíveis geradas por um modelo.
Amostragem	Uma técnica usada em modelos de linguagem onde um modelo gera respostas com base em uma distribuição de probabilidade, selecionando tokens aleatoriamente de acordo com suas probabilidades.
Pipeline de análise de sentimentos	Uma sequência de etapas de processamento na Hugging Face que avalia o sentimento (positivo ou negativo) de um texto, frequentemente usada para pontuar a qualidade das respostas geradas em modelos como chatbots.
Pontuação de sentimento	Uma pontuação que reflete o sentimento (positivo ou negativo) de uma resposta gerada, frequentemente usada no treinamento de modelos como PPO para incentivar a geração de respostas com um sentimento desejado.
Função sigmoide	Uma função matemática que produz uma curva em forma de S, comumente usada em aprendizado de máquina como uma função de ativação ou em regressão logística para mapear previsões em probabilidades.
Função softmax	Uma função usada em aprendizado de máquina para converter a saída de um modelo em uma distribuição de probabilidade, frequentemente aplicada na camada final de redes neurais para tarefas de classificação.
Stats_all	Uma lista ou armazenamento que contém as estatísticas de treinamento para cada lote em uma sessão de treinamento de otimização de política proximal (PPO), usada para rastrear métricas de desempenho.
Ascensão de gradiente estocástico (SGA)	Um algoritmo de otimização que atualiza iterativamente os parâmetros para maximizar uma função, particularmente útil em aprendizado por reforço onde o objetivo é maximizar recompensas esperadas.
Temperatura (τ)	Um hiperparâmetro na função softmax que controla a aleatoriedade das previsões escalando os logits antes de aplicar o softmax. Temperaturas mais baixas tornam a distribuição mais aguda, enquanto temperaturas mais altas a tornam mais uniforme.
Amostragem top-k	Um método em modelos de linguagem que restringe a seleção do próximo token aos k tokens de maior probabilidade, garantindo saídas mais focadas e coerentes ao filtrar opções menos prováveis.
Amostragem top-p	Uma técnica de amostragem onde o modelo seleciona o próximo token do menor conjunto de tokens cuja probabilidade cumulativa excede um limite p, permitindo uma amostragem mais dinâmica e dependente do contexto em comparação com a amostragem top-k.
Trainer.train()	Um método na classe Trainer da Hugging Face que inicia o processo de treinamento de um modelo usando o conjunto de dados e argumentos de treinamento especificados.
trainer.state	Uma propriedade na classe Trainer da Hugging Face que armazena o estado do processo de treinamento, incluindo registros de treinamento, que podem ser usados para monitorar o progresso e o desempenho do modelo durante o treinamento.

Registro de Mudanças

Data	Versão	Alterado por	Descrição da Mudança
2024-09-03	0.1	Palak J	Versão inicial criada

