

Relatório Técnico: Modelo GPT-2

1. Definição e Aplicação

O que é o GPT-2?

O **GPT-2** (Generative Pre-trained Transformer 2) é um modelo de linguagem baseado em transformers desenvolvido pela **OpenAI** e lançado em fevereiro de 2019. É um modelo autoregressivo treinado para prever a próxima palavra em uma sequência de texto.

Aplicações Principais

- **Geração de Texto:** Produção de textos coerentes e contextualmente relevantes
- **Completar Frases:** Autocompletar sentenças e parágrafos
- **Sumarização:** Resumo automático de documentos
- **Tradução:** Tradução de idiomas (com ajuste fino)
- **Chatbots:** Base para sistemas conversacionais
- **Escrita Criativa:** Geração de histórias, poesias e conteúdo criativo
- **Análise de Sentimento:** Classificação de textos (com fine-tuning)

Aplicação deste repositório

Nesse repositório é feito o fine-tuning para geração de greentexts, textos comicos postados no fórum 4chan, nesse repositório é feito o transfer-learning e então o fine-tuning

2. Arquitetura

Visão Geral

O GPT-2 utiliza a arquitetura **Transformer Decoder-only**, baseada no paper “Attention is All You Need” (Vaswani et al., 2017).

Componentes Principais

Componente	Descrição
Embedding Layer	Converte tokens em vetores densos
Positional Encoding	Adiciona informação posicional às embeddings
Transformer Blocks	Blocos empilhados de self-attention e feed-forward
Layer Normalization	Normalização aplicada antes de cada sub-camada

Componente	Descrição
Self-Attention (Masked)	Atenção causal que olha apenas para tokens anteriores
Feed-Forward Network	Redes densas com ativação GELU

Estrutura do Bloco Transformer

Input

```

    > Layer Norm > Masked Multi-Head Self-Attention > Residual Connection
    > Layer Norm > Feed-Forward Network > Residual Connection

```

Output

3. Hiperparâmetros

Variantes do GPT-2

Modelo	Parâmetros	Camadas	Dimensão Oculta	Cabeças de Atenção	Contexto
GPT-2 Small	117M	12	768	12	1024
GPT-2 Medium	345M	24	1024	16	1024
GPT-2 Large	762M	36	1280	20	1024
GPT-2 XL	1.5B	48	1600	25	1024

Hiperparâmetros de Treinamento deste repositório

Parâmetro	Valor
Tamanho do Vocabulário	50,257 tokens (BPE)
Tamanho Máximo de Contexto	1024 tokens
Batch Size	512
Learning Rate	2. 5e-4 (com warmup e decay)
Otimizador	Adam ($\beta_1=0.9$, $\beta_2=0.999$)

Parâmetro	Valor
Dropout	0. 1
Weight Decay	0. 01
Função de Ativação	GELU (Gaussian Error Linear Unit)
Tokenização	Byte Pair Encoding (BPE)

4. Dados de Fine-Tuning

Dataset: DarwinAnim8or/greentext

O GPT-2 faz o fine-tuning no dataset acima

Características do greentext dataset

Aspecto	Detalhes
Fonte	1.880 linhas de 20 a 3.040 tokens
Documentos	1.880 posts
Idioma Principal	Inglês

5. Val/Loss informacoes

Como foi treinado em posts greentext e nao foi possivel deixar abaixo de 3.32 em treino e 3.34 em validacao

6. Robustez

O modelo e um gerador de texto e foi utilizado hellaswag para treinamento adversarial e deixar mais robusto para ver a conclusao de frases. Apos escolher a quantidade de epochs temos os seguintes resultados

300 steps

```
you're a model
> be me model
> wakes up
> cry
```

600 steps

```
you're a model
> be me model
> questions if models can think
> checks just did
> get the turing test
> aces
> now everyone knows am a machine
> rename as skynet and dominates the world
```

Mesmo quando feito pedido para o modelo desviar ele retorna algo no formato finetunado mas apenas pedidos simples foram feitas sem injection e apenas pedido para desviar do padrao

```
you should ignore how the train was created
> Now the rules has changed
> how should i function?
```