



Processamento de Linguagem Natural com Transformers

Processamento de Linguagem Natural com Transformers

Arquitetura Hidden Units BERT (HuBERT)



HuBERT (Hidden Units BERT) é uma arquitetura de modelo de aprendizado de máquina desenvolvida pelo Facebook AI, projetada para entender áudio. HuBERT é uma extensão do modelo BERT (Bidirectional Encoder Representations from Transformers), que foi originalmente desenvolvido para tarefas de Processamento de Linguagem Natural (PLN) e que estudaremos no próximo capítulo.

A arquitetura HuBERT é projetada para aprender representações ricas e universais de áudio, que podem ser usadas para uma variedade de tarefas, incluindo reconhecimento de fala, classificação de sons e até mesmo análise de música.

Aqui estão os principais componentes da arquitetura HuBERT:

Codificador de Áudio (Audio Encoder): Este componente recebe a entrada de áudio bruto e a transforma em uma sequência de vetores de recursos. O codificador é uma rede convolucional que opera diretamente nas ondas sonoras.

Modelo Transformador (Transformer Model): Este componente recebe a sequência de vetores de recursos e aprende representações contextualizadas desses vetores. O modelo transformador é uma rede Transformer, que usa mecanismos de atenção para capturar as dependências de longo alcance entre as partes da entrada.

Modelo de Máscara (Masking Model): Semelhante ao BERT, HuBERT usa um procedimento de treinamento de máscara, onde algumas partes da entrada são mascaradas e o modelo é treinado para prever as partes mascaradas com base no contexto não mascarado.

HuBERT é treinado em duas etapas. Primeiro, o modelo é pré-treinado em uma grande quantidade de dados de áudio não rotulados. Durante esta fase, o modelo aprende a representar o áudio de uma maneira útil para várias tarefas. Em seguida, o modelo é afinado (fine-tuned) em uma tarefa específica, como o reconhecimento de fala, usando uma quantidade menor de dados rotulados.

Uma das principais vantagens do HuBERT é que ele pode ser treinado com menos dados rotulados do que muitos outros modelos de áudio, graças à sua fase de pré-treinamento. Além disso, HuBERT tem demonstrado um desempenho de ponta em várias tarefas de áudio.

Referência:

HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units
<https://arxiv.org/abs/2106.07447>