Transformers

This transformer library made by hugging face. when we use this transformer library. then you will be able to use the large language models, which is present inside the hugging face.

Hugging-face has also a datasets library. Jab hum datasets ko install karte hain, to ye hugging face ke library se. waha ke datasets ko use kar sakte hain. apne ~~machine~~ machine learning/LLM ko training ya fine-tune ya kisi down-streamed task ke liye train karna ho.

Sentencepiece → Thise library is made by google
ye tokenization me kaam aata hai

'Huggingface_hub → ye bhi ek library hai. Jo hugging-face provide karta hai. taaki aap kisi bhi NLP models ko access kar sako

Task for Text classification

```
from transformers import pipeline
classifier = pipeline ('text-classification')
```

→ Humlog ess tasko ko karne ke liye transformers se pipeline class ko import karenge. Jaise hi hum ess class ko import karke lein, to eske saath Jitni bhi eski functionality hai wo already eshe saath hamare system environment me import hu jaati hai.

agar hum pipeline ke ander ye define kar diya ('text classification') → to pipeline automatically ess task ko trigger kar dega.

aap pipeline me model = 'model name' bhi pass karte ho agar koi specific model aap import karna chahte ho

Pad = token_id 10

let's suppose hamare pass 2 ~~words~~ sentences hain.
(i) He is running
(ii) narendra-modi is the primeminister of india.

Dono sentence ka size alag-alag hai jab dono ko tokenize karenge to size alag-alag aayega. Tokens ke size ko equal karne ke liye padding kiya jata hai. esme hum kuchh no. add kar dete hain. taaki dono ka size equal ho sake. yaha par hum zero add kar rahe hain.

NER (Im use of Transformer based model)

hum yaha par pipeline class ko use karte hain usme hum 'ner'. pass kar dete hain, ek. parameter hai, (aggregation_strategy = 'simple') pass kar dete hain.

Evaluation matrices for ~~BERT~~ LLm (BERT)

(i) **Bert score :-**
Bert generated text ke quality ko evaluate karta hai

How → ?
esme hum prediction (generated text) ko apne kisi text se compare karte hain.

(ii) **Bleu :-** (Billinguil Evaluation Understudy)

mainly ye machine translation ke evaluation roe use kiya jata hai.

e.g. let's suppose humre koi model banaya to machine translation kar raha hai. yaha par bleu_score ye batayega ki original_language se next language me jo esne convert kiya hai. ye kitna sahi kiya hai.

matlab ek human kise translate karta aur ek human kise translate kar raha hai usko compare karta hai.

(2-2)

GLUE (General language Understanding evaluation):-

↳ LLm me GLUE measure karne ka matlab hai, ki hamare model ku general languages ki kitni jaankari hai

Perplexility:→

perplexity bhi ek evaluation matrix hai, perplexility se hame ye pata chalta hai ki hamara LLm model next word predict karne me kitna accurate hua. (perplexility score)

→ Agar perplexility ka score bahut kam aa raha hai; to model ka ~~scanor~~ performance bahut best mana jata hai

Llama is a model
langchain is a frameWork
Hugging face is a platform