

Hugging Face GPT

Author: Matthew Yeseta

HuggingFaces implementation of [HuggingGPT](#). HuggingGPT is use to connect LLMs (ChatGPT) with the open datasets on the community of Hugging Face.

This HuggingFacesGPT is using the following public dataset:

AdamLucek/apple-environmental-report-QA-retrieval

Set Imports

```
In [1]: from transformers import load_tool
import os
import torch
from transformers import GPT2Tokenizer, TFAutoModelForCausalLM
from datasets import load_dataset
import tensorflow as tf
```

WARNING:tensorflow:From C:\Users\matth\anaconda3\Lib\site-packages\keras\src\losses.py:2976: The name tf.losses.sparse_softmax_cross_entropy is deprecated. Please use tf.compat.v1.losses.sparse_softmax_cross_entropy instead.

```
In [2]: ## Setup OpenAPI and OpenAI API KEY
```

```
In [3]: import openai
from transformers import pipeline
openai.api_key = os.getenv("OPENAI_API_KEY")
```

C:\Users\matth\anaconda3\Lib\site-packages\transformers\utils\generic.py:260: FutureWarning: `torch.utils.pytree._register_pytree_node` is deprecated. Please use `torch.utils.pytree.register_pytree_node` instead.
torch.utils.pytree._register_pytree_node(

Setup model and HuggingGPT

Questions: how to start a community meeting to get people involved

Questions: what is the best approach to talk and introduce sustainability

```
In [5]: import tensorflow as tf
from transformers import TFGPT2LMHeadModel, GPT2Tokenizer, AutoTokenizer
from datasets import load_dataset

tf.get_logger().setLevel('ERROR')

try:
    dataset = load_dataset("AdamLucek/apple-environmental-report-QA-retrieval", split="train")
except ValueError as e:
    print(f"Error loading dataset: {e}")

model = TFAutoModelForCausalLM.from_pretrained("gpt2")
tokenizer = AutoTokenizer.from_pretrained("gpt2")
tokenizer.pad_token = tokenizer.eos_token

print("Available columns:", dataset.column_names)
```

All PyTorch model weights were used when initializing TFGPT2LMHeadModel.

All the weights of TFGPT2LMHeadModel were initialized from the PyTorch model.

If your task is similar to the task the model of the checkpoint was trained on, you can already use TFGPT2LMHeadModel for predictions without further training.

Available columns: ['question', 'chunk']

```
In [6]: while True:
    user_input = input("Ask question: ")

    if user_input.lower() == "exit":
        print("Ending the conversation.")
        break

    example = dataset[0]
    context = example["question"]
    prompt = f"Context: {context}\nEnvironmental Question: {user_input}"

    inputs = tokenizer(prompt, return_tensors="tf", padding=True, truncation=True)

    # Generate text using the model
    outputs = model.generate(
        inputs.input_ids,
        attention_mask=inputs.attention_mask,
        max_length=100,
        do_sample=True, # Corrected the typo here
```

```

top_p=0.9, # Use nucleus sampling to avoid only choosing the most likely tokens
no_repeat_ngram_size=2, # Avoid repeating phrases
pad_token_id=tokenizer.eos_token_id
)

generated_text = tokenizer.decode(outputs[0], skip_special_tokens=True)
print()
print(f"GPT-2: {generated_text}\n")
print()

```

Ask question: what is the best approach to talk and introduce sustainability

GPT-2: Context: What types of workshops does a business offer for its suppliers to promote sustainability?

Environmental Question: what is the best approach to talk and introduce sustainability to a consumer product in a competitive market? Can it be changed by a different company?

Nigel W. Dickson: In some areas, we have a long term problem with the sustainability question. Because there are so many variables, the way we define a sustainable product and market is not the most clear-cut way. So,

Ask question: how to start a community meeting to get people involved

GPT-2: Context: What types of workshops does a business offer for its suppliers to promote sustainability?

Environmental Question: how to start a community meeting to get people involved in our work and about our objectives? Where does the training to take place and how much do you pay for each session? How do we determine if it's worth it for our business? What is the most important aspect to include in any program or course? Is there a value in giving away the work for free for others? If so,

Ask question: exit

Ending the conversation.

```

In [7]: from transformers import pipeline

generator = pipeline("text-generation", model="gpt2")

def chat_with_agent():
    print()
    print("ChatGPT-like agent (Type 'exit' to stop the conversation)")

    while True:
        # Take user input
        user_input = input("User: ")

```

```

if user_input.lower() == "exit":
    print("Ending the conversation.")
    break

# Generate response from GPT-2
response = generator(user_input, max_length=100, num_return_sequences=1)

# Extract and print the generated text
generated_text = response[0]['generated_text']
print(f"GPT-2: {generated_text}\n")
print()

```

Chat with an Agent on continuous conversation

In [8]: chat_with_agent()

ChatGPT-like agent (Type 'exit' to stop the conversation)

User: what is so important for people to know on environment

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

GPT-2: what is so important for people to know on environment?" asked David Wurm, the director of the National Water Resources Council for Vermont, when the plan was unveiled at a recent workshop for people with water issues.

The Water Protectors are working together to end pollution by making sure Vermont's waterways are clean. More than 30 organizations support the Vermont Act, an initiative signed into law by then-Gov. Chris Burlington in 2016 that guarantees public, private and nonprofit lands for renewable energy, water

User: what did Vermont waterways use to clean pollution

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

GPT-2: what did Vermont waterways use to clean pollution? To make this matter even more urgent, the U.S. Supreme Court is set to do just that again in February.

So if you've ever wondered what the Supreme Court says about what water should be treated for, don't hesitate to visit their Web site. You'll see their page devoted to the subject of what constitutes clean water.

Advertisement

What is dirty water, what are we now allowed to use without a permit

User: how best to introduce sustainability to business

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

GPT-2: how best to introduce sustainability to business is to give us the tools to create the best business practices to get involved.

"As such, we are looking for sustainable companies and we seek to help them grow with a passion for energy, sustainability, and low-carbon solutions."

Eligible companies on the list include:

The Power Grid Industry (PGI) of Africa (which includes power from South Africa, India, West Africa and Zimbabwe);

(which includes power

User: exit

Ending the conversation.

In []: