Advanced Fine Tuning Drug Classification

Gangzhaorige Li

https://github.com/gangzhaorige/ML-OPENAi-CustomerSupport/tree/main/Fine-Tuning/2000DrugExamples

Project Overview Process

1. Read data from Excel Spreadsheet.
2. Convert data into dataset. Drug → Class
3. Use the dataset to fine tune the model.
4. Test the fine tuned model.

We will be fine tuning gpt3.5 turbo.

Read Data from Excel Spreadsheet

Requirement: pandas

```
df = pd.read_excel(
    'Medicine_description.xlsx',
    sheet_name='Sheet1',
    header=0,
    nrows=2000
)
```

Convert data into dataset. Drug→Class. Generating JSONL file.

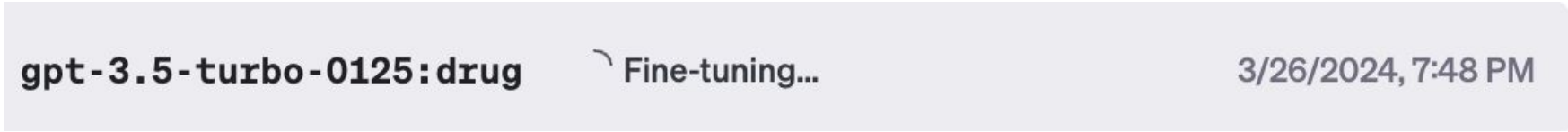
```python
# Get unique reasons
reasons = df["Reason"].unique()
# Create a dictionary to map reasons to indices
reasons_dict = {reason: i for i, reason in enumerate(reasons)}
df["Reason"] = df["Reason"].apply(lambda x: "" + str(reasons_dict[x]))
# Add system message
system_message = {"role": "system", "content": "You are profesional drug classifier."}
# Initialize the list to store messages
messages = []
# Iterate over rows in the DataFrame
for _, row in df.iterrows():
    # Add user message
    user_message = {"role": "user", "content": 'Drug: ' + row["Drug_Name"] + '.'}

    # Add assistant message
    assistant_message = {"role": "assistant", "content": row["Reason"]}

    # Construct a message group and add it to the list of messages
    message_group = [system_message, user_message, assistant_message]
    messages.append({"messages": message_group})

with open("output.jsonl", "w") as f:
    for msg in messages:
        json.dump(msg, f)
        f.write("\n")
```

JSONL file structure for GPT3.5Turbo

```
{
    "messages": [
        {"role": "system", "content": "You are profesional drug classifier."},
        {"role": "user", "content": "Drug: A CN Gel(Topical) 20gmA CN Soap 75gm."},
        {"role": "assistant", "content": "0"}
    ]
}
```

Use dataset to fine tune the Model

1. Go to https://platform.openai.com/finetune
2. Upload your jsonl file and choose your model.
3. Start fine tuning.



Testing Fine Tuned Model

1. After training is done. Model name would be provided.
2. Use that model to test it.

```
drugs = [
    "A CN Gel(Topical) 20gmA CN Soap 75gm",
    "Addnok Tablet 20'S",
    "ABICET M Tablet 10's",
]
```

Res: 0,5,2