

ChatGPT Applications – Project

Ngày 21 tháng 5 năm 2023

Phần I: Nội dung



Hình 1: ChatGPT API

ChatGPT (Chat Generative Pre-Training Transformer) là mô hình chatbot được phát triển bởi OpenAI ra mắt vào tháng 11 năm 2022. ChatGPT được xây dựng dựa trên việc tinh chỉnh mô hình ngôn ngữ lớn (Large Language Model) kết hợp với học tăng cường dựa trên feedback của con người (Reinforcement Learning with Human Feedback).

ChatGPT có thể giải quyết nhiều bài toán khác nhau trong xử lý tự nhiên như: phân loại văn bản (Text Classification), bài toán dịch máy (Machine Translation), bài toán hỏi đáp (Question Answering), bài toán tóm tắt văn bản (Text Summarization),... Từ các bài toán gốc cơ bản này, ChatGPT được triển khai cho nhiều lĩnh vực khác nhau: giáo dục, sáng tạo nội dung, chăm sóc sức khỏe,... Hơn thế nữa, một trong những bài toán điển hình mà ChatGPT có thể giải quyết là bài toán suy luận, giải toán.

Mô hình ChatGPT còn có khả năng tương thích, tối ưu cho bài toán cụ thể trong quá trình dự đoán (Inference - đưa ra phản hồi cho các bài toán) thông qua phương pháp học tập trong ngữ cảnh (In-Context Learning). Phương pháp học tập trong ngữ cảnh được xây dựng thông qua các **prompt**, được hiểu là các dữ liệu đầu vào với mục tiêu hướng dẫn hoặc mô tả chi tiết cho bài toán với mục tiêu giúp ChatGPT có thể xác định được bài toán cần xử lý và tối ưu miền xử lý, nhờ đó ChatGPT có thể cho kết quả tốt hơn.

Có ba cài đặt để xây dựng các prompt như sau:

- **Zero-shot:** Đưa ra câu trả lời chỉ dựa vào mô tả của bài toán

Trong hình dưới, prompt được định nghĩa với 2 phần. Phần 1, mô tả của bài toán: bài toán dịch máy từ ngôn ngữ tiếng anh (English) sang tiếng Pháp (French). Phần 2, dữ liệu đầu vào (văn bản cần dịch): cheese.



Hình 2: Cài đặt Zero-shot

- One-shot: Đưa ra câu trả lời dựa vào mô tả bài toán và một ví dụ mẫu cho bài toán

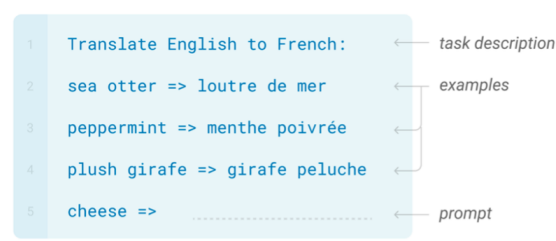
Trong hình dưới, prompt định nghĩa gồm 3 phần. Phần 1, mô tả của bài toán: bài toán dịch máy từ ngôn ngữ tiếng anh (English) sang tiếng Pháp (French). Phần 2, một ví dụ mẫu cho bài toán: sea otter sẽ được dịch thành loutre de mer. Phần 3, dữ liệu đầu vào (văn bản cần dịch): cheese.



Hình 3: Cài đặt One-shot

- One-shot: Đưa ra câu trả lời dựa vào mô tả bài toán và một vài ví dụ mẫu cho bài toán

Trong hình dưới, prompt định nghĩa gồm 3 phần. Phần 1, mô tả của bài toán: bài toán dịch máy từ ngôn ngữ tiếng anh (English) sang tiếng Pháp (French). Phần 2, một số các ví dụ mẫu cho bài toán: sea otter sẽ được dịch thành loutre de mer, peppermint được dịch thành menthe poivrée. Phần 3, dữ liệu đầu vào (văn bản cần dịch): cheese.



Hình 4: Cài đặt Few-shot

Các cài đặt ở trên và so với trường hợp mình không có mô tả nào và không có ví dụ mẫu cho bài toán thì tính hiệu quả của việc có mô tả và có ví dụ mẫu là vượt trội. Trong ba cài đặt ở trên thì Few-shot hiệu quả hơn One-shot và hiệu quả hơn Zero-shot.

Vì vậy, để tận dụng tốt ChatGPT cho các ứng dụng, bài toán khác nhau điều quan trọng chúng ta cần xây dựng các prompt phù hợp. Trong nội dung tiếp theo mình sẽ xây dựng một số prompt cho một số ứng dụng.

1. Sử dụng ChatGPT API

Để sử dụng được API của ChatGPT, chúng ta sẽ tạo tài khoản và lấy `OPENAI_API_KEY`, mô hình sử dụng là `text – davinci – 003`.

```
1 # Install library
2 !pip install --upgrade --quiet openai
3
4 # Import library
```

```

5 import openai
6 import IPython
7
8 # Set your OPENAI_API_KEY
9 OPENAI_API_KEY = "YOUR_OPENAI_API_KEY"
10
11 # OPENAI API configuration
12 openai.api_key = OPENAI_API_KEY
13
14 def set_open_params(
15     model="text-davinci-003",
16     temperature=0.7,
17     max_tokens=256,
18     top_p=1,
19     frequency_penalty=0,
20     presence_penalty=0,
21 ):
22     """ set openai parameters"""
23
24     openai_params = {}
25
26     openai_params['model'] = model
27     openai_params['temperature'] = temperature
28     openai_params['max_tokens'] = max_tokens
29     openai_params['top_p'] = top_p
30     openai_params['frequency_penalty'] = frequency_penalty
31     openai_params['presence_penalty'] = presence_penalty
32     return openai_params
33
34 def get_completion(params, prompt):
35     """ GET completion from openai api"""
36
37     response = openai.Completion.create(
38         engine = params['model'],
39         prompt = prompt,
40         temperature = params['temperature'],
41         max_tokens = params['max_tokens'],
42         top_p = params['top_p'],
43         frequency_penalty = params['frequency_penalty'],
44         presence_penalty = params['presence_penalty'],
45     )
46     return response
47
48 def print_response(response):
49     return IPython.display.Markdown(response.choices[0].text)
50
51 # Example:
52 # basic parameters
53 params = set_open_params()
54 prompt = " What is ChatGPT?"
55
56 response = get_completion(params, prompt)
57 print_response(response)

```

2. Sử dụng API ChatGPT thông qua thư viện langchain

```

1 # Install library
2 !pip install --upgrade --quiet openai langchain
3
4 # Import library
5 import openai

```

```

6 import os
7 import IPython
8 from langchain.llms import OpenAI
9 from langchain.chat_models import ChatOpenAI
10 from langchain import PromptTemplate, LLMChain
11 from langchain.schema import (
12     AIMessage,
13     HumanMessage,
14     SystemMessage
15 )
16
17 # Set your key
18 OPENAI_API_KEY = "YOUR_OPENAI_API_KEY"
19 os.environ["OPENAI_API_KEY"] = OPENAI_API_KEY
20 chat = ChatOpenAI(temperature=0)
21 USER_INPUT = "I love AI VIET NAM."
22 FINAL_PROMPT = """Classify the text into neutral, negative or positive.
23 Text: {user_input}.
24 Sentiment:"""
25
26 chat([HumanMessage(content=FINAL_PROMPT.format(user_input=USER_INPUT))])

```

3. Phân loại văn bản

Xét ví dụ, chúng ta không mô tả rõ yêu cầu cho bài toán phân tích cảm xúc thì ChatGPT sẽ phân loại nó dựa vào nhãn chung tổng quát.

```

1 prompt = """Classify the folowing text.
2 Text: I think the food was okay.
3 This text is:
4 """
5
6 response = get_completion(params, prompt)
7 print_response(response)
8 # => Opinion

```

Để ChatGPT tối ưu mục đích phân loại của bài toán thì chúng ta có 2 cách như sau:

- Cách 1: Thêm mô tả các nhãn vào phần mô tả bài toán

```

1 prompt = """Classify the folowing text into neutral, negative, postive.
2 Text: I think the food was okay.
3 This text is:
4 """
5
6 response = get_completion(params, prompt)
7 print_response(response)
8 # => Neutral

```

- Cách 2: Đưa thêm ví dụ để mô hình tự động hiểu được bài toán cần làm là gì

```

1 prompt = """Classify the folowing text into neutral, negative, postive.
2 Text: So bad.
3 This text is: negative.
4 Text: I think the food was okay.
5 This text is:
6 """
7
8 response = get_completion(params, prompt)
9 print_response(response)
10 # => Neutral

```

4. Bài toán dịch máy

Cũng tương tự như bài toán phân loại văn bản, chúng ta sẽ có hai cách cơ bản để tạo các prompt.

- Cách 1: Mô tả cho bài toán với ngôn ngữ đích vào ngôn ngữ nguồn. Tuy nhiên sẽ có kết quả trả về khác nhau giữa các prompt. Giải sử chúng ta cần dịch câu: Tôi đang học trí tuệ nhân tạo tại AI VIET NAM.

```

1 # Translation Prompt 1
2 prompt = """Translate these sentence from vi to en language.
3 Text:
4 """
5
6 response = get_completion(params, prompt)
7 print_response(response)
8 # => I am learning Artificial Intelligence at AI VIET NAM.
9
10 # Translation Prompt 2
11 prompt = """What do these sentences mean in en?
12 Text:
13 """
14
15 response = get_completion(params, prompt)
16 print_response(response)
17 # # => I am studying Artificial Intelligence at AI VIET NAM.
18
19 # Translation Prompt 3
20 prompt = """Please provide the en translation for these sentences:
21 Text:
22 """
23
24 response = get_completion(params, prompt)
25 print_response(response)
26 # => I am learning Artificial Intelligence at AI VIET NAM.
```

- Cách 2: Đưa thêm ví dụ để mô hình xác định ngôn ngữ nguồn và ngôn ngữ đích

```

1 prompt = """Translate these sentence.
2 Text: Toi thich no
3 Translated: I like it.
4 Text:
5 """
6
7 response = get_completion(params, prompt)
8 print_response(response)
9 # => I'm learning Artificial Intelligence at AI VIET NAM.
```

5. Bài toán hỏi đáp

Với bài toán hỏi đáp, chúng ta cần chú ý dữ liệu huấn luyện được thu thập vào trước năm 2021, vì vậy những câu hỏi có nội dung sau năm 2021 thì mô hình sẽ có thể trả lời sai hoặc không có đáp án phù hợp. Xét hai ví dụ sau

- Mô hình có thể trả lời những câu hỏi trước 2021

```

1 prompt = """Answer the following question:
2 Question: Where was President Ho Chi Minh born?
3 """
4
5 response = get_completion(params, prompt)
```

```

6 print_response(response)
7 # => Answer: Ho Chi Minh was born in 1890 in Hoang Tru, a village in central
  Vietnam.

```

- Với những câu hỏi sau 2021 chúng ta cần đưa thêm ngữ cảnh vào

```

1 # No context
2 prompt = """Answer the following question:
3 Question: How many gold medals does Vietnam have in SEA Games 2023?
4 """
5
6 response = get_completion(params, prompt)
7 print_response(response)
8 # => Answer: At this time, it is not possible to answer this question as the
  SEA Games 2023 have not taken place yet.
9
10 # With context
11 prompt = """Answer the following question:
12 Question: How many gold medals does Vietnam have in SEA Games 2023?
13 Context: Vietnam won 136 gold medals to secure the top spot in the medal
  tally as the 2023 Southeast Asian (SEA) Games in Cambodia came to the
  final official day of competition on Tuesday.
14 """
15
16 response = get_completion(params, prompt)
17 print_response(response)
18 # => Answer: Vietnam has 136 gold medals at the 2023 Southeast Asian (SEA)
  Games.

```

6. Bài toán đối thoại

Mô hình ChatGPT được tinh chỉnh và tối ưu cho dữ liệu các đoạn hội thoại. Vì vậy, khi chúng ta cung cấp đủ ngữ cảnh thì mô hình đưa ra kết quả tốt.

```

1 prompt = """This is a conversation between a customer and a polite, helpful
  customer service agent.
2 Customer: Hi, I'd like a refund for the coffee maker I ordered. Would that be
  possible?
3 AI: Response from the customer service agent: Hello. Thank you for reaching out
  to us. Yes...
4 """
5
6 response = get_completion(params, prompt)
7 print_response(response)
8 # => I'm sorry to hear that you're not satisfied with your purchase. We'll be
  glad to process a refund for you. Please provide us with the order number so
  we can expedite the process. Thank you for your time.

```

7. Bài toán suy luận

Ứng dụng khá thú vị và cũng là thách thức cho các bài toán xử lý ngôn ngữ tự nhiên hiện nay đó là bài toán suy luận. Tuy nhiên mô hình hiện tại xử lý tốt ở những tác vụ suy luận đơn giản, với những tác vụ suy luận phức tạp hơn thường cho kết quả không chính xác và không nhất quán trong việc suy luận ra kết quả.

- Mô hình có thể cho kết quả đúng với những suy luận đơn giản

```

1 prompt = """What is 9,000 * 9,000?
2 """
3

```

```

4 response = get_completion(params, prompt)
5 print_response(response)
6 # => 81,000,000

```

- Với những câu hỏi suy luận phức tạp hơn thì mô hình đưa ra các kết quả thường không nhất quán và không chính xác.

```

1 prompt = """The odd numbers in this group add up to an even number: 15, 32,
   5, 13, 82, 7, 1.
2 """
3
4 response = get_completion(params, prompt)
5 print_response(response)
6 # => The odd numbers in this group add up to 106, which is an even number.
7
8 prompt = """The odd numbers in this group add up to an even number: 15, 32,
   5, 13, 82, 7, 1.
9 """
10
11 response = get_completion(params, prompt)
12 print_response(response)
13 # => The odd numbers add up to 118, which is an even number.

```

- Trong trường hợp những bài suy luận phức tạp, chúng ta cần chỉ ra đúng các bước suy luận thì mô hình cho kết quả tốt.

```

1 prompt = """The odd numbers in this group add up to an even number: 15, 32,
   5, 13, 82, 7, 1.
2 Solve by breaking the problem into steps. First, identify the odd numbers,
   add them, and indicate whether the result is odd or even.
3 """
4
5 response = get_completion(params, prompt)
6 print_response(response)
7 # => Odd numbers: 15, 5, 13, 7, 1
8 # Sum of odd numbers: 41
9 # Result: 41 is an odd number.

```

8. Ứng dụng cho viết mã nguồn

ChatGPT được huấn luyện dữ liệu mã nguồn thu thập từ Github có thể viết code, sửa một số lỗi cơ bản, viết mô tả hoặc viết lại code rõ ràng hơn.

- Viết code, tuy nhiên vẫn có một số lỗi nhỏ như dùng tên trùng với từ khoá.

```

1 prompt = """Write a function that calculates the average of the numbers in
   Python List.
2 """
3
4 response = get_completion(params, prompt)
5 print(response.choices[0].text)
6 # result
7 def average(nums):
8     total = 0
9     for num in nums:
10         total += num
11     return total / len(nums)
12
13 prompt = """Write a function that calculates the average of the numbers in
   Python List.

```

```

14 """
15
16 response = get_completion(params, prompt)
17 print(response.choices[0].text)
18 # result
19 def averageList(list):
20     total = 0
21     for item in list:
22         total += item
23
24     return total/len(list)

```

- Viết mô tả cho code

```

1 prompt = """Write a function that calculates the average of the numbers in a
  Python List.
2 Add to comment for code
3 """
4
5 response = get_completion(params, prompt)
6 print(response.choices[0].text)
7 # result
8 def calculate_average(numbers):
9     """Calculate the average of a list of numbers.
10
11     Args:
12         numbers (list): List of numbers
13
14     Returns:
15         float: Average of the numbers
16     """
17     total = 0
18     for num in numbers:
19         total += num
20     return total / len(numbers)

```

- Gợi ý sửa lỗi code

```

1 prompt = """Debug this code:
2 def average(nums):
3     total = 0
4     for num in nums
5         total += num
6     return total/len(nums)
7 """
8
9 response = get_completion(params, prompt)
10 print(response.choices[0].text)
11 # result
12 #Fix
13 def average(nums):
14     total = 0
15     for num in nums: # Colon added
16         total += num
17     return total/len(nums)

```


Phần II: Trắc nghiệm

1. Phương pháp huấn luyện của mô hình ChatGPT là gì?
 - (a) Tinh chỉnh mô hình ngôn ngữ lớn GPT3.5 (Large Language Model)
 - (b) Tinh chỉnh mô hình ngôn ngữ lớn BLOOM
 - (c) Tinh chỉnh mô hình mBART
 - (d) Tinh chỉnh mô hình ngôn ngữ lớn GPT 3.5(Large Language Model) với phương pháp học tăng cường dựa trên phản hồi người dùng.
2. Bộ dữ liệu để tinh chỉnh mô hình ChatGPT là gì?
 - (a) Dữ liệu tóm tắt văn bản
 - (b) Dữ liệu là các cuộc hội thoại
 - (c) Dữ liệu phân loại văn bản
 - (d) Dữ liệu dịch máy
3. Phương pháp cài đặt nào không có trong phương pháp học trong ngữ cảnh (In-Content Learning) được sử dụng trong quá trình sinh văn bản là gì?
 - (a) One-Shot
 - (b) Few-Shot
 - (c) Fine-Tuning
 - (d) Zero-Shot
4. Bài toán nào sau đây ChatGPT có thể giải quyết?
 - (a) Sinh nhạc từ văn bản
 - (b) Sinh video từ văn bản
 - (c) Chuyển ảnh từ 2D sang ảnh 3D
 - (d) Sửa lỗi chính tả
5. Hạn chế của ChatGPT là gì?
 - (a) Không giải quyết được bài toán hỏi đáp
 - (b) Không giải quyết được bài toán trích xuất thông tin (Information Extraction)
 - (c) Không được cập nhật dữ liệu sau năm 2021
 - (d) Không giải quyết được bài toán trích xuất thực thể (Named Entity Recognition)
6. Thư viện nào sau đây hỗ trợ sử dụng ChatGPT API?
 - (a) Spacy
 - (b) Fasttext
 - (c) Langchain
 - (d) Fairseq

Quan sát cách cài đặt các prompt sau đây và trả lời câu hỏi 1 đến 5:

```
1 # Prompt 1
2 prompt = ""Classify the folowing text into neutral, negative, postive and
   explain why result is like this.
3 Text: I think the food was okay.
4 This text is:
5 ""
6
7 # Prompt 2
8 prompt = ""Classify the folowing text into neutral, negative, postive and
   explain why result is like this.
9 Text: So bad.
```

```

10 This text is: negative.
11 Text: I think the food was okay.
12 This text is:
13 ""
14
15 # Prompt 3
16 prompt = ""Classify the folowing text into neutral, negative, postive and
    explain why result is like this.
17 Text: So bad.
18 This text is: negative.
19 Text: I love it.
20 This text is: positive.
21 Text: I think the food was okay.
22 This text is:
23 ""
24
25 # Prompt 4
26 prompt = ""
27 Text: So bad.
28 This text is: negative.
29 Text: I love it.
30 This text is: positive.
31 Text: I don't like this food.
32 This text is: negative.
33 Text: I think the food was okay.
34 This text is:
35 ""

```

7. Prompt 1 sẽ là cài đặt nào trong các cài đặt sau đây dựa vào phương pháp học trong ngữ cảnh (In-Content Learning)?

(a) One-Shot	(c) Fine-Tuning
(b) Zero-Shot	(d) Few-Shot
8. Prompt 2 sẽ là cài đặt nào trong các cài đặt sau đây dựa vào phương pháp học trong ngữ cảnh (In-Content Learning)?

(a) One-Shot	(c) Fine-Tuning
(b) Zero-Shot	(d) Few-Shot
9. Prompt 3 sẽ là cài đặt nào trong các cài đặt sau đây dựa vào phương pháp học trong ngữ cảnh (In-Content Learning)?

(a) One-Shot	(c) Fine-Tuning
(b) Zero-Shot	(d) Few-Shot
10. Prompt 4 sẽ là cài đặt nào trong các cài đặt sau đây dựa vào phương pháp học trong ngữ cảnh (In-Content Learning)?

(a) One-Shot	(c) Fine-Tuning
(b) Zero-Shot	(d) Few-Shot
11. Dựa vào cách cài đặt ở prompt 1, để mô hình có thể giải thích ví do ví sao mô hình chọn nhãn cho văn bản đầu vào, prompt nào sau đây là phù hợp nhất?

- (a) Classify the following text into neutral, negative, positive and explain why result is like this.
- (b) Sentiment Analysis.
- (c) Translate English language to Vietnamese language
- (d) Classify the following text. For example: I love it => Positive

- *Hết* -