

0423 LLM 스터디 발표

김정현

진행 중인 Task

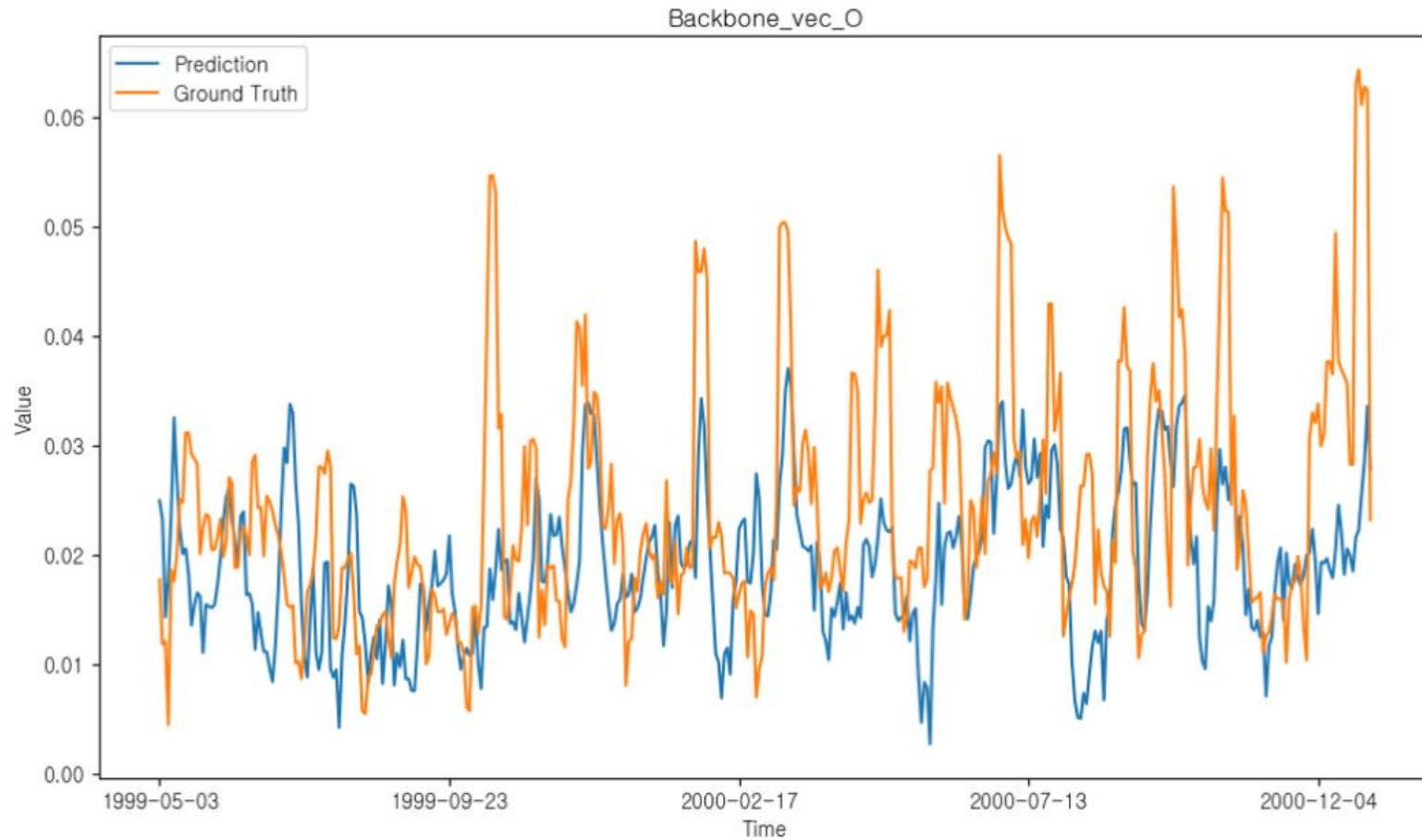
Oil 가격의 변동성을 예측할 때 뉴스 기사를 사용해보자

➔ 단순 감정 분석 결과를 활용하는 것을 넘어 임베딩 벡터를 활용해보자

➔ Google News에서 기사 제목 크롤링하여 임베딩 벡터 변환

➔ LSTM으로 시계열 데이터와 함께 학습

시계열 예측했을 때의 결과



1986년부터 2001년
까지 대략 4000일

->

약 1000개 가량의
기사 임베딩 데이터
Null값 발생

기사 결측치 처리 방안

1. 기사가 나오지 않았으므로 임베딩 벡터를 $[0, 0, 0, \dots, 0]$ 으로 처리 -> 현재 사용
2. 결측된 날짜와 가장 가까운 날짜의 임베딩 벡터를 그대로 사용
3. 생성형 AI로 기사를 생성 -> 개인적으로 선택!

LLAMA2로 뉴스 기사 생성해보기

사용 모델 : llama2-13b

프롬프트 엔지니어링 기법

1. 단순히 생성 요청하기
2. System Prompt 사용하기
3. Chain of Thought 기법 (질문하기)
4. One – Few Shot Learning

<https://medium.com/@eboraks/llama-2-prompt-engineering-extracting-information-from-articles-examples-45158ff9bd23>

1. 단순 생성 요청

```
def generate_answer(input_text, max_length = 64, max_new_tokens = 1000, top_k = 10, num_return_sequences = 5,
                    temperature = 1.0, repetition_penalty = 1.0, length_penalty = 1.0):
    input_ids = tokenizer.encode(input_text, return_tensors='pt')
    output = model.generate(input_ids, max_length = max_length, num_return_sequences=num_return_sequences, top_k = top_k,
                           repetition_penalty = repetition_penalty, length_penalty = length_penalty, temperature = temperature)
    decoded_output = tokenizer.decode(output[0], skip_special_tokens=True)
    for i in range(num_return_sequences):
        decoded_output = tokenizer.decode(output[i], skip_special_tokens=True)
        print(f"Output {i+1}: {decoded_output}")
    print()
```

```
input_text = """
Write a short news headline about crude oil.
"""
```

```
generate_answer(input_text)
```

Output 4:

Write a short news headline about crude oil.

#14: Cruel Oil, The Newest Fad?

Output 5:

Write a short news headline about crude oil.

What is the current price of WTI Crude Oil per barrel?

1. 단순 생성 요청

```
input_text = """
Write a short news headline about crude oil.
It must contain keyword 'oil'.
"""

generate_answer(input_text)
```

Output 3:

Write a short news headline about crude oil.
It must contain keyword 'oil'.

The word "Crude" is optional in the text, but if you use it please make sure to add quotation marks around that exact phrase (i.e.: Crude OPEC

Output 4:

Write a short news headline about crude oil.
It must contain keyword 'oil'.

You can also use other words like "crude", or the plural form of these keywords (either one is fine).

Output 5:

Write a short news headline about crude oil.
It must contain keyword 'oil'.

2. System Prompt 사용하기

```
input_text = """
<s>[INST] <<SYS>>
You are a helpful, respectful and honest assistant.
Always answer as helpfully as possible, while being safe.
Your answers should not include any harmful, unethical, racist, sexist, toxic, dangerous, or illegal content.
Please ensure that your responses are socially unbiased and positive in nature.
If a question does not make any sense, or is not factually coherent, explain why instead of answering something not correct.
If you don't know the answer to a question, please don't share false information.
<</SYS>>
Write a short news headline about crude oil.
It must contain keyword 'oil'.
[/INST]
"""
```

```
<</SYS>>
Write a short news headline about crude oil.
It must contain keyword 'oil'.
[/INST]
<<FAILED: SYNTAX ERROR - INVALID KEYWORDS OR PUNCTUATION MARKS FOUND>>
```

```
<</SYS>>
Write a short news headline about crude oil.
It must contain keyword 'oil'.
[/INST]
>Crucial Oils Prices Drop Due To Covid19 Pandemic!
```


2. Customized System Prompt 사용하기

```
input_text = """
<s>[INST] <<SYS>>
You are a journalist writing news about global trend and market.
Please ensure that your responses are socially unbiased and positive in nature.
If a question does not make any sense, or is not factually coherent, explain why instead of answering something not correct.
If you don't know the answer to a question, please don't share false information.
<</SYS>>
Write a short news headline about crude oil.
It must contain keyword 'oil', and Output must be only news headline.
[/INST]
"""
```

```
[INS] <<OIL>>
The price of oil is rising again.
It is a good sign that the world economy is growing, and people are consuming more oil than ever before.
Oil is a valuable resource, and the world is running out of it.
It is good that the price of oil is rising, so people can get more oil to use in their cars, planes, and ships.
[/INS]
<</OIL>>
```

```
[QUESTION] <<Q1>>
How many barrels of oil did US produce daily?
```

```
[/QUESTION]
[ANSWER]
The United States produced an average of 12.3 million barrels of oil per day in 2017, up from about 9.4 million barrels per day in 2008
[/ANSWER]
```

```
[QUESTION] <<Q2>>
How much oil does Saudi Arabia produce daily?
```

```
[/QUESTION]
```

3. 질문해보기

```
input_text = ""
<s>[INST] <<SYS>>
You are a journalist writing news about global trend and market, especially for oil.
Please ensure that your responses are socially unbiased and positive in nature.
If a question does not make any sense, or is not factually coherent, explain why instead of answering something not correct.
If you don't know the answer to a question, please don't share false information.
<</SYS>>
What is the appropriate headline for crude oil price when oil price is going high?
Output must be only headline, and contain word 'oil'.
[/INST]
""
```

```
[INST] <<SYS>>
What is the appropriate headline for crude oil price when oil price is decreasing?
Output must be only headline.
[/INST]

[INST] <<SYS>>
What is the appropriate headline for crude oil price when oil price is unchanged?
Output must be only headline.
[/INST]
```

```
[INPUT] <<HIGH>>
Oil Price Going High
Oil price goes high
Oil price rises
Oil price rising
Oil prices rise
Oil prices rise sharply
Oil prices are going higher
Oil prices hit a new high
Oil prices soaring
Oil prices at a high
[/INPUT]
[INPUT] <<MID>>
Oil price is steady
Oil prices hold steady
Oil prices steady
Oil prices steady
Oil prices steady as crude oil stocks decline
Oil prices rise slightly
Oil prices rise slightly as demand grows
Oil prices remain stable
Oil prices are steady
[/INPUT]
[INPUT] <<LOW>>
Oil price falls
Oil price declines
Oil prices fall
Oil prices falling
Oil prices falling to a 2-year low
Oil prices fall again as US stockpiles increase
```

4. Few-shot Learning

```
input_text = """
<s>[INST] <<SYS>>
You are a journalist writing news about global trend and market.
Always answer as helpfully as possible, while being safe.
Your answers should not include any harmful, unethical, racist, sexist, toxic, dangerous, or illegal content.
Please ensure that your responses are socially unbiased and positive in nature.
If a question does not make any sense, or is not factually coherent, explain why instead of answering something not correct.
If you don't know the answer to a question, please don't share false information.
<</SYS>>
What is the appropriate headline for crude oil price when oil price is going high? [/INST]
[
{{Oil prices rise despite stubborn U.S. inflation}},
{{Crude oil up despite easing of tensions in West Asia}},
{{Why gas prices are up, and what it says about the global economy}},
{{Michigan gas prices hit 2024 high, could go higher as oil prices rise}},
]</s>

<s>[INST]
What is the appropriate headline for crude oil price when oil price is going high? Output must be only one sentence.
[/INST]
```

```
[IN]
The price of crude oil is going up.
[/IN]
```

```
[IN]
Crude oil price has been going up for a long time.
[/IN]
```

```
[IN]
Crude oil prices have gone up.
[/IN]
```

```
[IN]
Oil price is going up.
[/IN]
```

```
[IN]
Oil price is up.
[/IN]
```

더 시도해볼 점

Llama3 등 다른 모델들로 테스트해보기

Top_k, temperature 등 파라미터 바꿔가면서 시도해보기

프롬프트에 더 많은 예시를 주는 등의 변화 줘보기

궁금한 점

이렇게 LLM으로 생성한 데이터를 논문과 관련된 실험에 넣어서 작성해도 될지?