구루의 생성형 언어모델 app 만들기

Supplementary content

임베딩(Embedding)이란?

지난 시간 요약 : Transformer는 왜 성능이 좋을까?

트랜스포머가 성능이 좋은 이유는

'attention' 이라는 매커니즘을 이용하여

정확히는
scaled dot-product attention

인풋 sequence (문장)의 길이에 상관없이

심도 깊은 context의 파악이 가능하므로!!

Attention Is All You Need

임베딩(Embedding) 이란?

@#%^&*()_+-
=[]{}|:';",.<>?/!~

사람

%$#*^!~|+=_[]{}<>:;/&@
()

LLM

니네 서로 말은 통하니?

임베딩(Embedding) 이란?

사람의 언어 ≠ LLM의 언어

임베딩 (embedding)

임베딩(Embedding) 이란?

사람의 언어

English

日本語

한국어

임베딩 (LLM의 언어)

GPT 임베딩

Llama 임베딩

Falcon 임베딩

임베딩(Embedding) 이란?

사람의 언어

- 문자열로 이루어져있다

안녕하세요

こんにちは

Hello

你好

- 각 언어는 문자와 문법이 다르다

임베딩 (LLM의 언어)

- 숫자열로 이루어져있다
  벡터

[0.0305, 0.6202, 0.0402, 0.8760]

[-0.8883, -0.8489, 0.9559, 0.9249, 0.8823, 1.5023]

[-0.1303, 1.4558, 2.1038, -0.8730, -0.0796, -1.6468, 1.6711, 0.1072]

- 각 임베딩은 벡터의 길이와 숫자를 정하는 규칙이 다르다

임베딩(Embedding) 이란?

사람

질문 by 사람의 언어

점심 메뉴 추천해줘

피자 어때요?

답변 by 사람의 언어

Embedding model

질문 by 임베딩

[[-0.1303, 1.4558, 2.1038, -0.8730, -0.0796, -1.6468, 1.6711, 0.1072, 1.1142, -0.9374] ... ...]

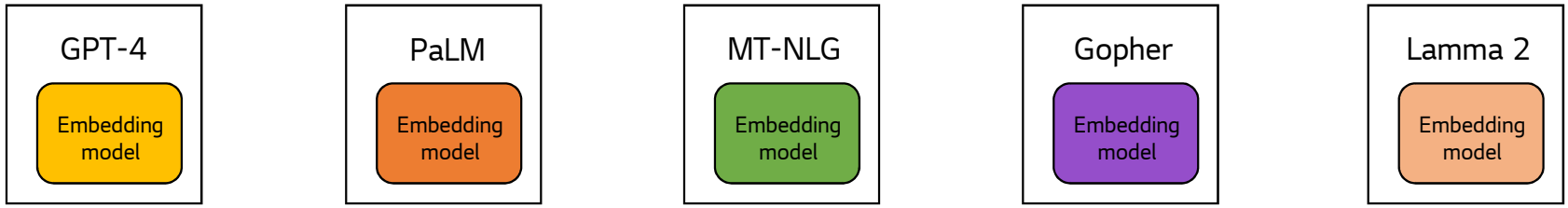
[[0.0305, 0.6202, 0.0402, 0.8760, -0.8883, -0.8489, 0.9559, 0.9249, 0.8823, 1.5023] ... ...]

답변 by 임베딩

LLM

임베딩 모델이 사람과 LLM 사이에서 동시통역사의 역할을 한 답니다

임베딩 (Embedding) 이란?



모든 LLM은 반드시 자신만의 동시통역사, 즉 임베딩 모델을 내장하고 있답니다.

임베딩 (Embedding) 이란?

I love you so much.

임베딩 모델 인풋 (인간의 언어)

Embedding model

<<시작>>	[0.4810, -0.7042, -0.7409, 0.14624]
I	[-0.5198, 0.2901, 0.3235, -0.5886]
love	[1.5240, 2.5387, -1.0701 , -0.1190]
you	[-0.2612 , 0.1227, -0.4248 , 0.6229]
so	[-0.7988, 1.6670, 0.0759, -1.2467]
much	[0.1693 , 1.7550, 0.3056, 0.0773]
.	[-0.0600, 0.9258, -1.2276, 0.4466]
<<끝>>	[1.3262 , -0.8511, 1.4349, -0.6320]

[[0.4810, -0.7042, -0.7409, 0.14624],
[-0.5198, 0.2901, 0.3235, -0.5886],
[1.5240, 2.5387, -1.0701 , -0.1190],
[-0.2612 , 0.1227, -0.4248 , 0.6229],
[-0.7988, 1.6670, 0.0759, -1.2467],
[0.1693 , 1.7550, 0.3056, 0.0773],
[-0.0600, 0.9258, -1.2276, 0.4466],
[1.3262 , -0.8511, 1.4349, -0.6320]]

임베딩 모델 아웃풋 (임베딩 행렬)

모든 단어를 길이 4인 벡터로 변환시키는 임베딩 모델의 예제입니다.

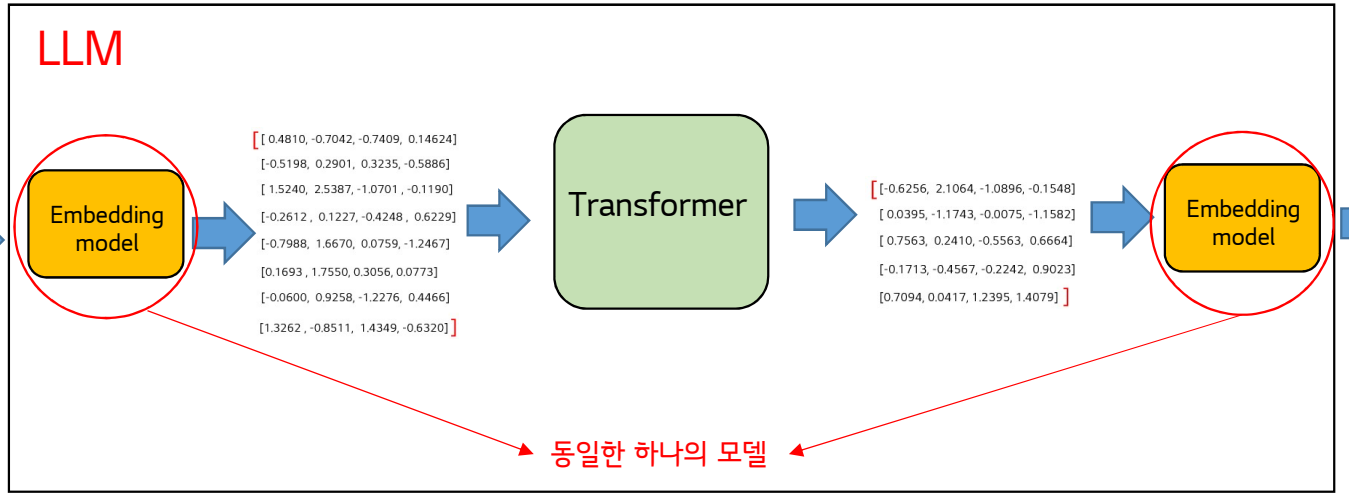
'시작' 플래그와 '끝' 플래그를 포함하여 모든 단어를 길이 4인 벡터로 변환하며, 각각의 벡터를 임베딩 또는 임베딩 벡터라고 합니다.

임베딩 벡터들은 프로세싱이 용이하도록 행렬(matrix) 형태로 반환됩니다. 이 아웃풋을 임베딩 혹은 임베딩 행렬이라고 합니다.

임베딩 (Embedding) 이란?

I love you so much.

LLM 인풋
(사람의 언어)

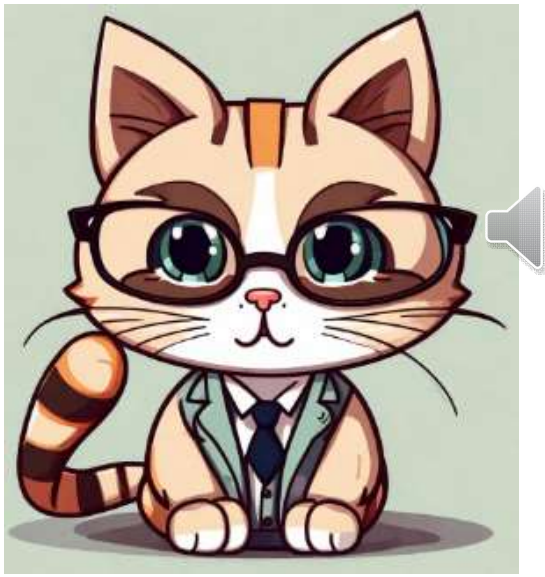


사랑해

LLM 아웃풋
(사람의 언어)

이해를 돕기 위해 두개의 임베딩 모델을 그렸지만 실제로 임베딩 모델은 하나입니다.

임베딩 (Embedding) 이란?



임베딩은 트랜스포머의 핵심인
scaled dot-product attention
(QKV attention)을 이해하기 위해
필요하답니다.
하지만 이정도만 알면 충분하니 걱정
마세요~