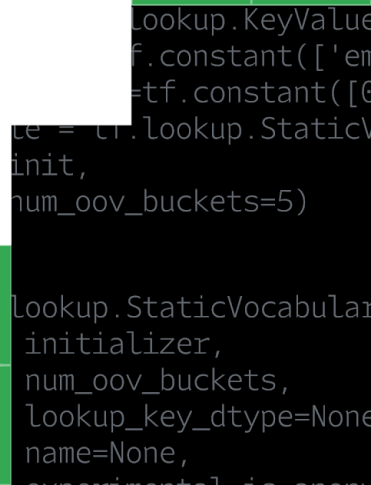
Google Developer Student Clubs

4주차 2조

팀원: 황동욱, 권도혁, 송성근



(황동욱) 알고리즘 정리

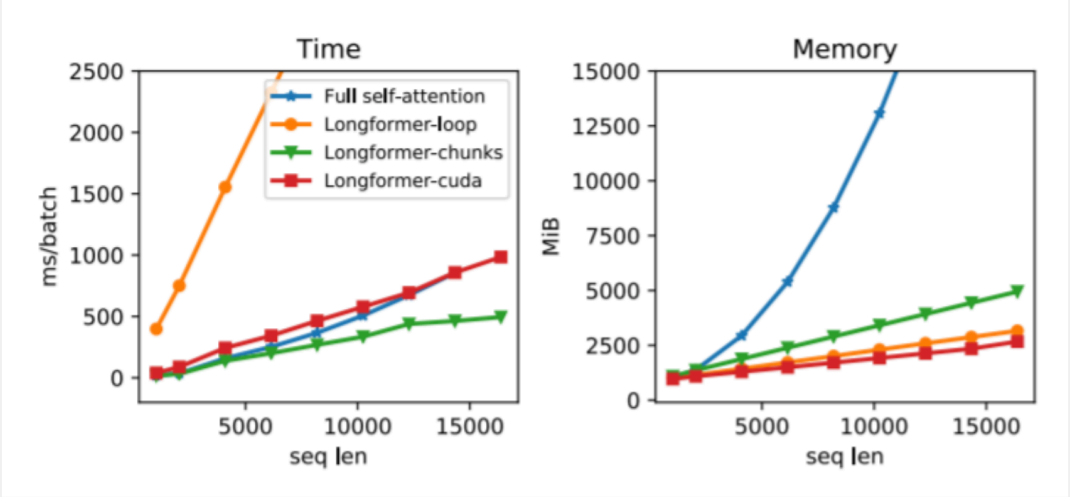
BERT vs Longformer attention 메커니즘

- **BERT**: BERT는 표준적인 self-attention 메커니즘을 사용합니다. 이는 입력 시퀀스의 모든 토큰 간의 상호 작용을 계산합니다. 따라서 BERT의 attention은 O(n^2)의 복잡도를 가집니다. 여기서 n은 시퀀스 길이입니다. 이러한 계산 복잡도 때문에 BERT는 긴 문서를 처리하는 데 어려움이 있습니다.
- **Longformer**: Longformer는 긴 문서를 처리할 수 있도록 설계된 attention 메커니즘을 사용합니다. 이는 global attention과 local attention의 조합으로 구성됩니다. 특정 중요한 토큰은 전체 문서에 걸쳐 attention을 받는 반면, 나머지 토큰은 주변의 토큰만을 대상으로 attention을 수행합니다. 이러한 방식으로 Longformer는 O(n)의 복잡도로 확장됩니다.

tokenizer와 vectorizer의 차이

tokenizer와 vectorizer의 차이:

- **tokenizer**: 토크나이저는 주로 딥러닝 모델에서 사용되며, 텍스트를 모델이 처리할 수 있는 형태로 변환하는 역할을 합니다. 특히, 토크나이저는 텍스트를 개별 토큰(단어나 부분 단어)으로 분리하고, 각 토큰을 해당 토큰의 고유한 ID로 매핑하는 역할을 합니다. 여기서 사용된 `LongformerTokenizer`는 Longformer 모델에 특화된 토크나이저입니다.
- **vectorizer**: 벡터라이저는 텍스트를 수치 벡터로 변환하는 역할을 합니다. `TfidfVectorizer`는 텍스트를 TF-IDF 값으로 변환하는 벡터라이저입니다. TF-IDF는 텍스트 내의 각 단어의 중요도를 나타내는 값으로, 단어의 빈도와 문서 내에서의 등장 빈도를 기반으로 계산됩니다.



Longformer와 Transformer의 메모리 소모량 비교

(권도혁) 알고리즘 정리

Longformer의 Motivation

기존 Transformer의 문제점?

- Transformer의 self-attention은 입력 시퀀스의 길이 제곱에 비례하여 메모리 및 계산량이 증가하는데, 이로 인해 이전에는 긴 텍스트를 처리하는 자연어 이해 작업에서 성능 제한이 발생
- BERT는 이를 해결하기 위해 입력 문서의 일부를 삭제하여 512 글자 제한에 맞추고 Stride를 사용하여 독립적인 문서 조각들을 처리하며 관련된 문서 조각을 추출하고 두 단계의 추론을 통해 모델을 실행함.

따라서 Longformer는

- 시퀀스 길이에 선형적으로 비례하는 메모리 및 계산량 스케일의 attention 메커니즘을 제안.
- 기존의 self-attention을 즉시 대체할 수 있는(drop-in 방식으로 대체 가능) 방법을 제시.

이를 통해 기계독해, 문서 분류, 일관성 모델링과 같은 다양한 NLU 작업에서 실험을 수행함.

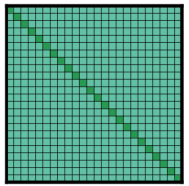
Longformer에서 사용한 Model 들

1. Sliding Window

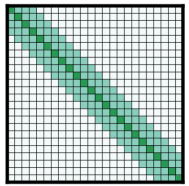
- 주변 토큰들과 고정 크기의 윈도우 내에서만 관심을 두며, 윈도우 밖에 있는 토큰들에 대해서는 고려하지 않는 방식. 이러한 레이어를 여러 층으로 쌓을 경우, 넓은 문맥 정보를 담고 있는 표현을 생성할 수 있음

- 윈도우 방식의 패턴에서는 각 단어 주변의 토큰 중 윈도우 사이즈의 절반만큼만 주의를 기울이기 때문에, 입력 시퀀스의 길이 n에 대한 선형 계산 복잡도를 가짐.

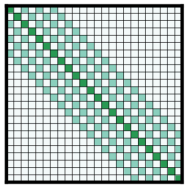
- 즉, 윈도우 크기가 작을수록 효율적으로 작동.



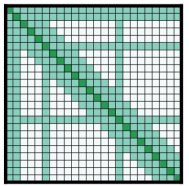
(a) Full n^2 attention



(b) Sliding window attention



(c) Dilated sliding window



(d) Global+sliding window

Figure 2: Comparing the full self-attention pattern and the configuration of attention patterns in our Longformer.

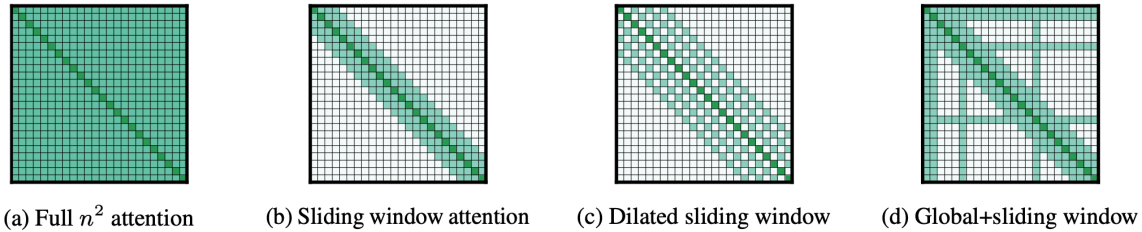
Longformer에서 사용한 Model 들

2. Dilated Sliding Window

- Sliding Window와 유사하지만, 토큰을 건너뛰는 방식으로 윈도우를 확장하여 구성. 건너뛰는 토큰 수를 작게 설정하더라도 기본 윈도우 방식보다 훨씬 넓은 영역에 대한 표현을 생성할 수 있음.

- 각 head 마다 건너뛰는 토큰 수(dilation size)를 다르게 적용하여, 어떤 head는 지역적인 정보에 집중하고, 다른 head는 더 넓은 문맥 정보에 집중할 수 있도록 함.

- 결과적으로 각 head가 다양한 문맥에 집중할 수 있도록 함.



(a) Full n^2 attention (b) Sliding window attention (c) Dilated sliding window (d) Global+sliding window

Figure 2: Comparing the full self-attention pattern and the configuration of attention patterns in our Longformer.

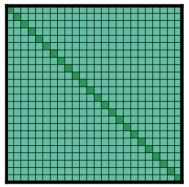
Longformer에서 사용한 Model 들

3. Global window

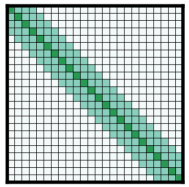
- 텍스트 분류 및 질의응답과 같은 작업에서는 [CLS], [SEP]와 같은 특별한 토큰을 사용하는데, 이러한 토큰은 문장 전체적으로 attention을 적용하여 학습하는 것이 중요. 따라서 특별한 토큰과 다른 토큰에 global attention을 적용.

- 이러한 방식은 특별한 토큰의 개수가 적어 global attention을 적용해도 계산의 양이 많지 않음에도, 문맥적인 표현을 학습할 수 있음.

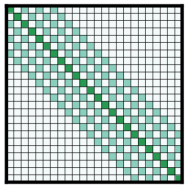
- 기존 QKV의 가중치를 학습하는 것과 비슷한 효과이나, Longformer에서는 이를 분리하여 적용함.



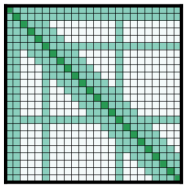
(a) Full n^2 attention



(b) Sliding window attention



(c) Dilated sliding window



(d) Global+sliding window

Figure 2: Comparing the full self-attention pattern and the configuration of attention patterns in our Longformer.

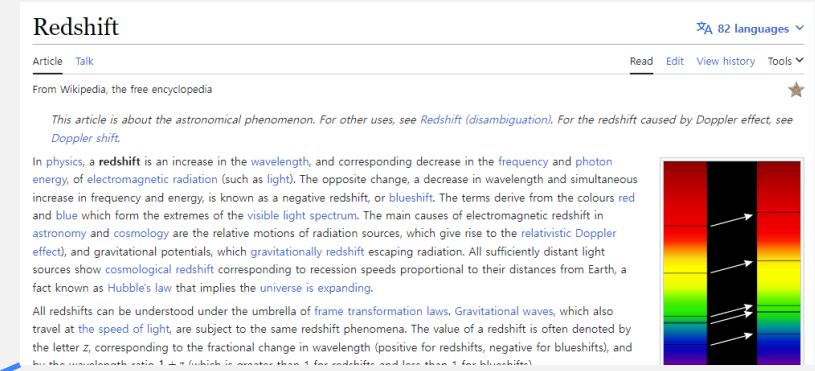
(송성근) 알고리즘 정리

Open Book LLM

문제와 관련된 문서를 Wikipedia에서 찾아 Context로 문제, 답안과 함께 Tokenize를 해준다.

What is the term used in astrophysics to describe light-matter interactions resulting in energy shifts in the radiation field?

A. Blueshifting
B. Redshifting
C. Reddening
D. Whitening
E. Yellowing



Context

Prompt

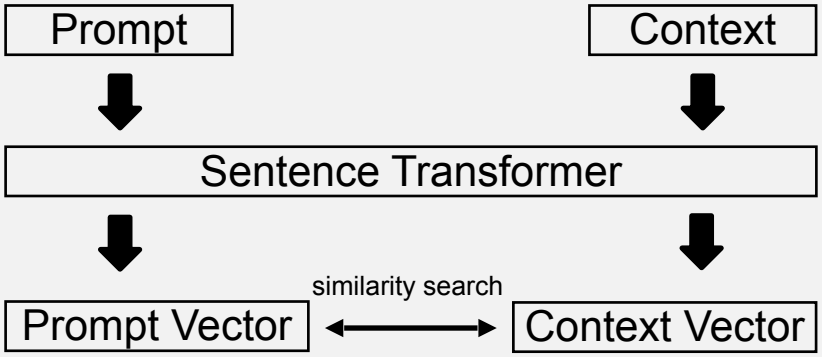
Answer

Faiss

Faiss는 Facebook에서 개발한 Vector들의 유사도를 구하는 라이브러리이다.

Prompt와 Context를 Sentence Transformer를 통해 Embedding하고
이를 Faiss를 이용해 가장 비슷한 Prompt, Context 쌍을 구한다.

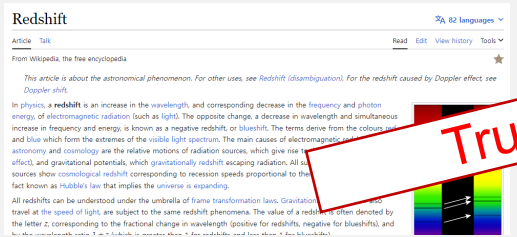
이를 활용하여 Prompt를 통해 Wikipedia에서 가장 비슷한 문서를 Token으로 사용한다.



DeBERTa-V3 Max Token

BERT 모델은 Input으로 입력받는 최대 토큰의 개수를 512개로 설정되어 있고 이를 초과할 시 일정 토큰을 설정한 정책에 따라 일부 버리게 된다.

따라서 Context, Prompt, Answer가 모두 Tokenize 되어야 하는 이번 과제의 특성상 높은 정확도를 가질 수 없었다고 예상된다.


Redshift

82 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

This article is about the astronomical phenomenon. For other uses, see Redshift (disambiguation). For the redshift caused by Doppler effect, see Doppler shift.

In physics, a **redshift** is an increase in the wavelength, and corresponding decrease in the frequency and photon energy, of electromagnetic radiation (such as light). The opposite change, a decrease in wavelength and simultaneous increase in frequency and energy, is known as a negative redshift, or blueshift. The terms derive from the colours red and blue which form the extremes of the visible light spectrum. The main causes of electromagnetic redshift in astronomy and cosmology are the relative motions of radiation sources, which give rise to the Doppler effect, and gravitational potentials, which gravitationally redshift escaping radiation. All sufficiently distant sources show cosmological redshift corresponding to recession speeds proportional to their distances, a fact known as Hubble's law that implies the universe is expanding.

All redshifts can be understood under the umbrella of frame transformation laws. Gravitational redshifts, which travel at the speed of light, are subject to the same redshift phenomena. The value of a redshift is often denoted by the letter *z*, corresponding to the fractional change in wavelength (positive for redshifts, negative for blueshifts), and

Truncated

What is [illegible]sed in astrophysics to [illegible]t-matter interactions resulting [illegible]energy shifts in the radiation field?

A. Blueshifting
B. Redshifting
C. Reddening
D. Whitening
E. Yellowing

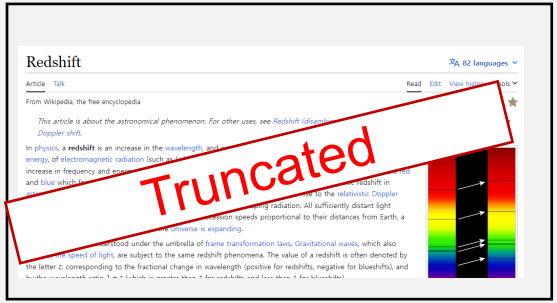
BOS | Context | Prompt | EOS | BOS | Answer | EOS > 512

DeBERTa-V3 Max Token 개선?

기존 모델을 Finetuning 하는 비용이 너무 커 Kaggle에서 다른 유저들이 학습시킨 모델을 이용했다.

위 모델들은 앞서 페이지와 같이 Context와 Prompt를 하나의 Sentence로 묶은 후 Answer과 함께 Tokenize를 진행하였는데 이 때 Context보다 더 높은 중요도를 가지는 Prompt가 Truncated될 수 있다.

만약 Context와 Prompt를 떨어뜨린 후 Tokenize를 진행하고 이를 활용하여 Finetuning을 하게 된다면 상대적으로 덜 중요한 Context가 Truncated되어 더 높은 정확도를 가질 수 있지 않을까 예상된다.



What is the term used in astrophysics to describe light-matter interactions resulting in energy shifts in the radiation field?

A. Blueshifting
B. Redshifting
C. Reddening
D. Whitening
E. Yellowing

BOS | Context | EOS | BOS | Prompt | EOS | BOS | Answer | EOS > 512