

SBERT는 BERT의 문장 임베딩의 성능을 우수하게 개선시킨 모델이다.

SBERT를 학습하는 방법은 2가지가 존재

NLI(Natural Language Inferencing) 문제

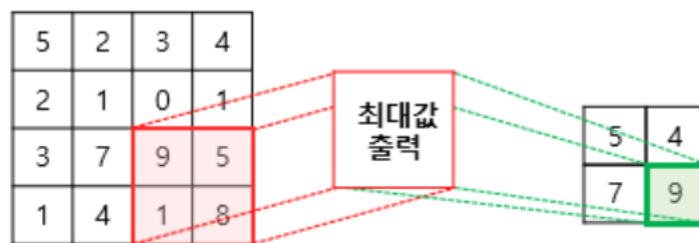
- NLI는 두 개의 문장이 주어지만 수반 관계인지, 모순 관계인지, 중립관계인지 맞추는 문제
- SBERT는 NLI 문제를 학습하기 위해 다음과 같은 과정을 거친다.
 - 문장 A와 B를 BERT의 입력으로 넣고, 평균 pooling 또는 MAX pooling을 통해 각각에 대한 문장 임베딩 벡터를 얻는다.
 - 그 후, 두 개의 문장에서 얻은 벡터의 차이를 구한 벡터를 구한다.
 - 문장 A에서 얻은 벡터, B에서 얻은 벡터, 차이를 구한 벡터를 연결한다. (세 개의 벡터를 하나의 벡터로 연결)
 - 연결한 벡터를 출력층으로 보내 다중 클래스 분류 문제를 푼다. (수반, 모순, 중립 문제 해결)

STS(Semantic Textual Similarity) 문제

- STS는 두 개의 문장으로부터 의미적 유사성을 구하는 문제이다.
- 레이블은 두 문장의 유사도로 범위 값은 0~5(실수)이다.
- SBERT는 STS 문제를 학습하기 위해 다음과 같은 과정을 거친다.
 - 문장 A와 문장 B를 BERT의 입력으로 넣고 평균 pooling, MAX pooling을 통해 각각에 대한 문장 임베딩 벡터를 얻는다.
 - 두 개의 문장 임베딩 벡터를 통해 코사인 유사도를 구한다.
 - 해당 유사도와 레이블 유사도와의 평균 제곱 오차(MSE)를 최소화하는 방식으로 학습
 - 이때, 코사인 유사도의 값의 범위는 -1~1 사이이다.
 - 만약 레이블의 범위가 0~5이면 학습 전 레이블 값들을 5로 나누어 값의 범위를 줄여야한다.

Pooling

- 풀링은 특정 맵을 다운 샘플링하여 특정 맵의 크기를 줄이는 연산이다.
 - 4x4 행렬을 2x2 행렬로 줄이는 과정으로 생각
- 풀링연산을 할 때 커널의 크기가 2x2이면 맵에서 2x2 범위에서 최대 값을 뽑거나 평균 값을 뽑는 방식으로 연산한다.
- 풀링 연산은 가중치가 없으며 연산 후 채널 수가 변하지 않는다.



Pooling을 진행할 때 평균 풀링을 한다면 해당 문장 벡터는 모든 단어의 의미를 반영하는 쪽에 가깝다.

하지만 맥스 풀링을 한다면 문장 벡터는 중요한 단어의 의미를 반영하는 쪽에 가깝게 된다.