

# [논문리뷰]Simple Data Transformations for Mitigating the Syntactic Similarity to Improve Sentence Embeddings at Supervised Contrastive Learning 24 Sep 2024

논문-한양대(최영석교수님) · 2024. 9. 26. 19:45

[논문리뷰]Simple Data Transformations for Mitigating the Syntactic Similarity to Improve Sentence Embeddings at Supervised Contrastive Learning 24 Sep 2024

## •Contrastive Learning

(가까운것은 가깝게 먼것은 멀게)은 다양한 NLP 작업들에서 매우 좋은 성능을 보여주었음

•예를 들자면, Contrastive Learning을 통해 문장 임베딩을 학습하고 이것을 글 분류나 자료 검색 등에 사용하는 것

하지만, 이러한 학습 과정에서 모델은 실제 주어진 문장들을 이해하는 것이 아닌, 단순 패턴에 집중하는 듯한 모습을 보여줌

예를 들자면, 문장 속 단어의 순서들이 바뀌어도 동일한 결과를 도출하였음

이는 주어진 Contrastive Learning 데이터셋(NLI)들의 문장 구조들이 매우 비슷하기 때문임  
다 비슷한 문장들만 있다 보니, 모델이 문장 구조에 대한 학습을 제대로 하지 못한 것 (다양한 예시를 보지 못함)

따라서 모델은 문장 구조를 중요시하지 않고, 단순 단어의 분포 등에 의존하였음(예..이단어는 이렇게 예측)

또한 한편으로는, **문장 구조에 과하게 의존하는 경향도 보임** (문장 **구조 비슷하면 높은 유사도**, **아니면 낮은 유사도**)

•이 문제를 극복하기 위해, 본 연구에서는

1. **주어진 데이터셋의 문장들이 얼마나 구조적으로 비슷한지 평가하는 방법** 가져와서 **비교**
2. 주어진 문장에 대해, **의미를 유지하며 문장 구조를 바꾸는 효율적인 방법을 제시하고 (선행 연구)**, 이를 적용했을 때 **성능 향상 폭**을 측정

-**구문 트리 (Syntactic tree)**는 어떠한 문장의 **구조를 트리 형태로** 나타낸 것을 말함

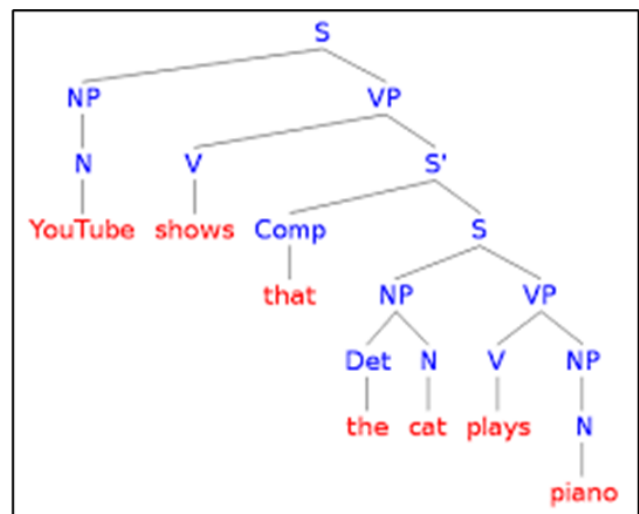
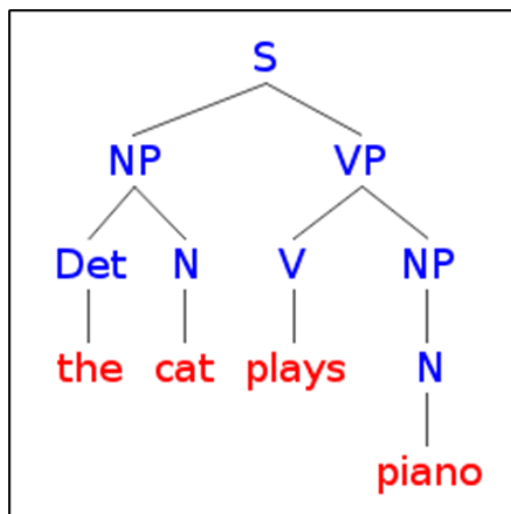
-일반적으로 다른 문장은 다른 구문 트리를 가지고 있지만, 반대의 경우도 가능함

-다른문장이 같은 구문트리를

-가질수있고,

-하나의 문장이(애매해서)

-사람마다생각하는 구문트리가 다를수있다



♡ 1



구독하기

자연어(NLP)

네이쳐2024 님의 블로그입니다.



구독하기 +

댓글 1



익명



비밀댓글입니다.

2024. 9. 26. 20:09



이름

비밀번호

내용을 입력하세요.



등록