

# Week 1-1 과제 : 6팀\_정태호


## 1. 본인이 본 강의를 수강하는 목적에 대해서 자유롭게 적어주세요

NLP 부트캠프를 통하여 자연어 처리에 대한 기술을 배우고 심화 시킬 수 있을 것으로 기대하여 지원하게 되었습니다.

실무와 유사한 환경을 경험해보고 팀원과 함께 과제를 해결해 나가는 경험을 통해 업무 능력을 발전 시켜보고 싶습니다.

## 2. Paperswithcode(<https://paperswithcode.com/area/natural-language-processing>)에서 NLP sub task 중에 2개를 선택하여 본인 블로그에 정리해보세요. task 별로 아래 3가지 항목에 대해서 정리하세요.

- 아이디 : infinite13
- URL : <https://blog.naver.com/infinite13/222653560303>

<div><div>◆task : Sentence Pair Modeling</div><div>◆subtask : Semantic Similarity</div><div><a href="https://paperswithcode.com/task/semantic-similarity">https://paperswithcode.com/task/semantic-similarity</a></div><div><div><div><div>Papers with Code - Semantic Similarity</div><div>The main objective of Semantic Similarity is to...</div><div><a href="https://paperswithcode.com">paperswithcode.com</a></div></div></div></div><div><div>▶문제 정의</div><div>단어, 문장 또는 문서에서의 의미적 유사도를 측정하는 것입니다. 단어간의 의미적 차이를 학습시키는 문제로 나아가 문장의 유사도를 비교하여 문장간의 논리적(감정적) 차이를 계산해내는 것입니다. 이를 통해 단순한 장르선에서의 유사도를 넘어 글의 문맥, 진행방식, 감정적 요소와 같은 복합적인 요인을 포함하는 분류 및 탐색 문제를 해결할 수 있습니다.</div><div><div>▶데이터 소개</div><div>sick(Sentences Involving Compositional Knowledge)로 구성적 분포 의미론에 대한 데이터 세트입니다. 문장 A,B가 존재하고 두 문장의 관계도를 1~5로 표시해 두었으며 두 문장을 유사,중립,모순을 구분하는 라벨링 데이터로 구성되어있습니다. 또한 문장 A 다음 B문장이 오는것 혹은 문장 B 다음 A문장이 오는 것이 자연스러운지에 대한 라벨링이 존재합니다. <a href="https://huggingface.co/datasets/sick">https://huggingface.co/datasets/sick</a></div><div><div>sick · Datasets at Hugging Face</div><div>Dataset Preview Go to dataset viewer Subset Split id (string) sent...</div><div><a href="https://huggingface.co">huggingface.co</a></div></div><div><div>▶SOTA : SimCSE-RoBERTalarge</div><div>SimCSE 메소드를 사용할때 드롭아웃을 통한 노이즈로 데이터증강이 발생하며 이를 통한 unsupervised 학습으로 BERT기반 모델에서는 유의미한 성능 향상을 보여 줍니다. <a href="https://velog.io/@seopbo/SimCSE-Simple-Contrastive-Learning-of-Sentence-Embeddings">https://velog.io/@seopbo/SimCSE-Simple-Contrastive-Learning-of-Sentence-Embeddings</a></div><div><div>SimCSE- Simple Contrastive Learning of Sentence Embe...</div><div>본 포스트는 BERT, RoBERTa와 같은 masked language model(MLM)...</div><div><a href="https://velog.io">velog.io</a></div></div><div>주요 키워드 : SimCSE</div></div></div></div></div>	<div><div>◆task :Speech Recognition</div><div><a href="https://paperswithcode.com/task/speech-recognition">https://paperswithcode.com/task/speech-recognition</a></div><div><div><div>Papers with Code - Speech Recognition</div><div>Speech recognition is the task of recognising speech within audi...</div><div><a href="https://paperswithcode.com">paperswithcode.com</a></div></div></div><div><div>▶문제 정의</div><div>음성인식으로 오디오 내에서 음성을 인식하고 텍스트로 변환 하는 문제입니다. 다양한 상황속에서 음성을 분리해 내고 텍스트로 변환하여 자연어 처리가 가능하게 합니다.</div><div><div>▶데이터 소개</div><div>LibriSpeech는 LibriVox 프로젝트의 일부인 약 1,000시간의 오디오북 모음입니다. 데이터는 클린 및 기타 범주로 분할되며 오디오의 길이는 5시간입니다. 또한 803M의 토큰과 977K의 고유단어를 포함하는 n-그램 모델과 해당 텍스트를 제공합니다. <a href="https://paperswithcode.com/dataset/librispeech">https://paperswithcode.com/dataset/librispeech</a></div><div><div>Papers with Code - LibriSpeech Dataset</div><div>The LibriSpeech corpus is a collection of approximately 1,000 ho...</div><div><a href="https://paperswithcode.com">paperswithcode.com</a></div></div><div><div>▶SOTA : Conformer + Wav2vec 2.0 + SpecAugment-based Noisy Student Training with Libri-Ligh</div><div>높은 성능의 모델을 만들기 위해 준지도 학습의 최근 방법을 사용하여 Wav2vec 2.0 사전훈련 모델을 기반으로 높은 성능을 달성하였습니다.</div></div></div></div></div>
--	---

### 3. 팀원들의 게시글을 읽고 피드백 댓글을 달아보세요

- 아이디 / 닉네임 : infinite13 / infinite13
- URL : [https://ossified-heat-913.notion.site/WEEK-01-01-\\_bd465059d57344f6ac0fe0582919b40b](https://ossified-heat-913.notion.site/WEEK-01-01-_bd465059d57344f6ac0fe0582919b40b)

## [WEEK-01-01] 정태호\_피드백

👤 작성자

태호 정

🕒 속성

February 21, 2022 4:32 PM

태호 Add a comment...

### 손지아

task에 대한 정보를 간략하게 잘 작성하신 것 같습니다.

데이터 셋에 대한 정보를 이미지로 참조해두셔서 한눈에 확 들어와서 좋습니다.

### 최진수

내용을 간략히 잘 정리하셔서 작성하신 것 같아요.

중요한 내용을 강조해두셔서 읽기 수월했습니다.

### 현승환

내용을 간략히 잘 설명해 주셨지만 문단 구분이나 글자 크기가 다 일정해서 내용을 읽기가 불편해진 것 같습니다.

정태호 → 팀원 게시물에 대해 작성한 피드백 댓글 내용입니다. (노션)