



문장 유형 분류 AI 경진대회

제출수 늘려주세요



개발환경

OS

- windows 10
- ubuntu 22.04
- aws ec2 p3 2xlarge

라이브러리

- torch 1.13.0
- cuda 11.7.1

성능을 올리기 위한 시도

1. Text Augmentation

- replaced 2 words in one sentence and made 2 or 3 sentences.(randomly)

2. Ensemble

3. Customize model layers

- Residual Blocks

4. Change loss function

- Focal Loss
- Asymmetric Loss

(https://github.com/Alibaba-MIIL/ASL/blob/main/src/loss_functions/losses.py)

(<https://paperswithcode.com/paper/asymmetric-loss-for-multi-label>)

5. Fold(just 5 folds)

6. Huggingface Custom Trainer 🙌

7. Custom Scheduler (안 썼습니다.)

앙상블 경로

- 740
 - .ipynb
- 741
 - .ipynb
- 743
 - .ipynb
- 744
 - .ipynb
- 749
 - .ipynb
- 7474
 - .ipynb
- 앙상블.ipynb

각 폴더의 숫자는 추론을 제출했을때의 **Public Score** 입니다.

사용한 사전학습모델과 출처

- **kykim/electra-kor-base**

(<https://github.com/kiyoungkim1/LMkor>)

(<https://huggingface.co/kykim/electra-kor-base>)

(<https://openreview.net/pdf?id=r1xMH1BtvB>)

- **monologg/kobigbird-bert-base**

(<https://github.com/monologg/KoBigBird>)

(<https://huggingface.co/monologg/kobigbird-bert-base>)

(<https://arxiv.org/abs/2007.14062>)

- **beomi/KcELECTRA-base-v2022**

(<https://github.com/Beomi/KcELECTRA>)

(<https://huggingface.co/beomi/KcELECTRA-base-v2022>)

(<https://openreview.net/pdf?id=r1xMH1BtvB>)

자세한 사항은 소스를
봐주세요 !

감사합니다