

문장 유형 분류 AI 경진대회

600마력 팀

puzzle_collector, 왕만두박사, Gmin47

SUMMARY

- **[Public 10th 0.75615/ Private 2nd, 0.75769]**

- 모델

- Monologg/kobigbird-bert-base
- 참고논문: <https://arxiv.org/abs/2007.14062>

- 개발환경

- OS

- Windows 10
- Ubuntu 18.04

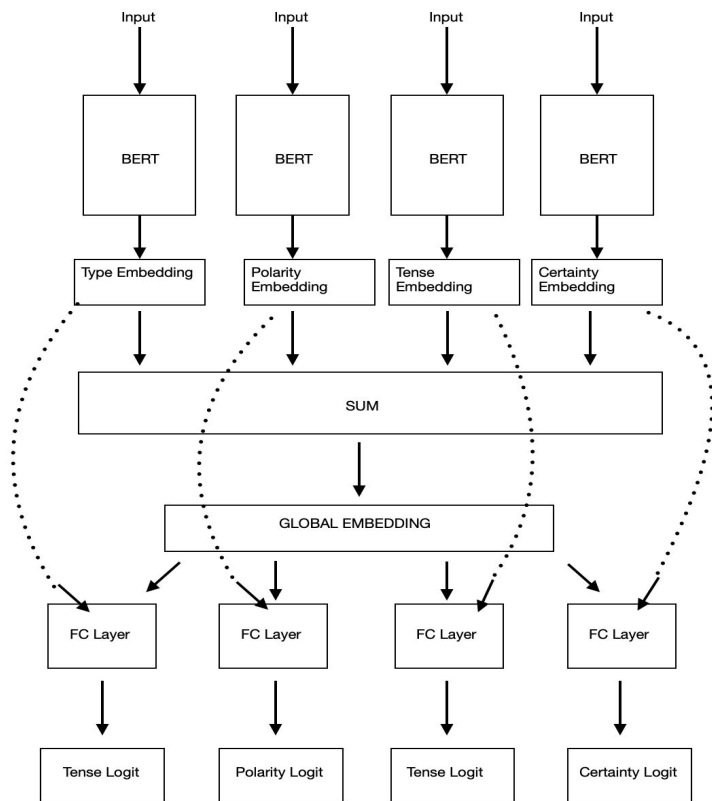
- GPU

- NVIDIA A100-80GB

- 라이브러리

- Torch 1.9
- Cuda 11.1
- transformers 4.22.1
- iterative-stratification 0.1.7

Model Architecture



- 그림의 구조와 같이 4개의 언어모델을 활용하여 각 타겟 임베딩 학습을 수행함
- 실험 결과 성능이 좋은 mean pooling을 활용함
- 최종 FC Layer에 추가로 Multi Sample Dropout을 적용

PERFORMANCE STRATEGY

- Hyperparameter setting
 - CrossEntropyLoss
 - Batch size : 16
 - Optimizer
 - AdamW
 - Learning rate : $2e-5$
 - no warmup & linear decay scheduling
- Fold
 - 10-fold cross validation (iterative stratification)
- Ensemble
 - 각 Fold로 검증했을때 가장 validation F1 score 가 높은 체크포인트 10개를 이용해서 soft voting 진행
- Data Augmentation (미적용)
 - Back-Translation (Korean – English – Korean) 시도해봤으나 성능 개선에는 실패

FILE PATH

- 파일 경로

- [학습 및 추론 파일]
 - ./BigBirdX4_Daemon_submission.py
 - ./infer_bigbirdx4.py
- [데이터]
 - ./data/{train, test, submission}.csv

- 사전학습 모델 출처

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

Thank you