

# 문장 내 개체간 관계 추출 2등 Solution

---

## 쿼터백 조

김다영 \* 김다인 \* 박성호 \* 박재형 \* 서동건 \* 정민지 \* 최석민

# 목차

---

- Input Format
- Data Augmentation
- Model Architecture
- Ensemble

# Input Format

SENTENCE : 이순신은 조선시대 중기 무신이다.

이순신 [SEP] 무신 [SEP] 이순신은 조선시대 중기 무신이다. [SEP]



Add Entity Marker

[SBJ]이순신[/SBJ] [OBJ]무신[/OBJ] [SEP] + SENTENCE with special token

[SBJ+TYPE]이순신[/SUB] [SEP] [OBJ+TYPE]무신[/OBJ] [SEP] + SENTENCE with special token



Add Query

[SBJ+TYPE]이순신[/SUB]과 [OBJ+TYPE]무신[/OBJ]의 관계 [SEP] + SENTENCE with special token

[SBJ+TYPE]이순신[/SUB]과 [OBJ+TYPE]무신[/OBJ]의 관계는? [SEP] + SENTENCE with special token

[SBJ+TYPE]이순신[/SUB]과 [OBJ+TYPE]무신[/OBJ]의 관계는 무엇일까? [SEP] + SENTENCE with special token

# Input Format

Method	Input Example
Entity mask	[ SUBJ-PERSON] was born in [OBJ-CITY].
Entity marker	[E1] Bill [/E1] was born in [E2] Seattle [/E2].
Entity marker (punct)	@ Bill @ was born in # Seattle #.
Typed entity marker	<S:PERSON> Bill </S:PERSON> was born in <O:CITY> Seattle </O:CITY>.
Typed entity marker (punct)	@ * person * Bill @ was born in # ^ city ^ Seattle #.

Zhou, Wenxuan, and Muhao Chen. "An Improved Baseline for Sentence-level Relation Extraction. 2021

Example sentence : 이순신은 조선시대 중기 무신이다.

Subject entity: 이순신

Object entity: 무신

} → Query: 이순신과 무신의 관계

<Input>

[CLS] @ \* PER \* 이순신 @ 과 # ^ POH ^ 무신 # 의 관계 [SEP] @ \* PER \* 이순신 @ 은 조선시대 중기 # ^ POH ^ 무신 # 이다 [SEP]

# Data Augmentation – SBJ & OBJ Entity Random Masking

[CLS] @ \* PER \* 이순신 @ 과 # ^ POH ^ 무신 # 의 관계 [SEP] @ \* **PER \* 이순신** @ 은 조선시대 중기 # ^ POH ^ 무신 #이다 [SEP]



SBJ & OBJ Entity  
Random Masking

Subject  
Masking

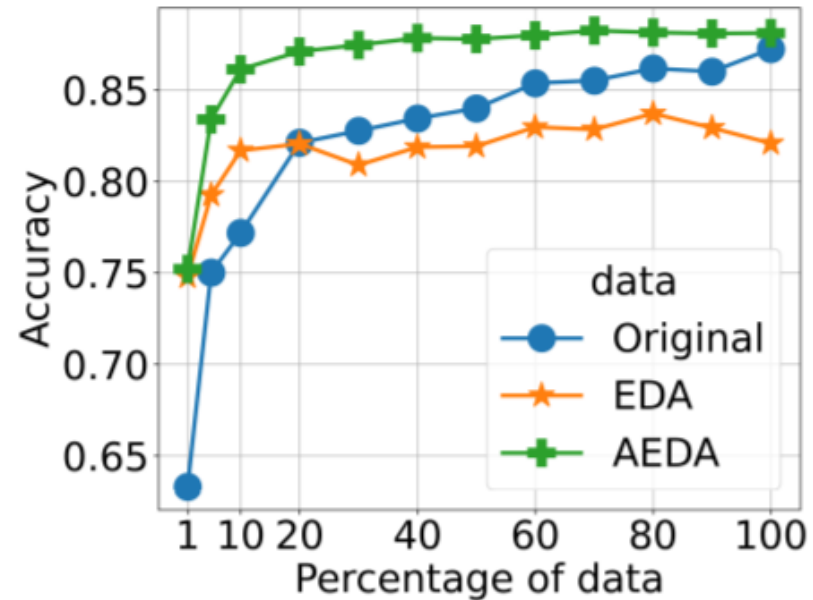
[CLS] @ \* PER \* 이순신 @ 과 # ^ POH ^ 무신 # 의 관계 [SEP] @ \* **PER \* [MASK]** @ 은 조선시대 중기 # ^ POH ^ 무신 #이다 [SEP]

Object  
Masking

[CLS] @ \* PER \* 이순신 @ 과 # ^ POH ^ 무신 # 의 관계 [SEP] @ \* PER \* 이순신 @ 은 조선시대 중기 # ^ POH ^ **[MASK]** #이다 [SEP]

50% 확률로 Subject Entity Masking, 50% 확률로 Object Entity Masking

# Data Augmentation – AEDA



Karimi, Akbar, Leonardo Rossi, and Andrea Prati. "AEDA: An Easier Data Augmentation Technique for Text Classification.". 2021

".", ",", "!", "?", ";", ":"

를 Okt 형태소 분석기 기준 토큰나이징한 결과에 특정 확률로 삽입  
기존 문장 + AEDA로 생성한 문장 사용

@ \* PER \* 이순신 @ 은 조선시대 중기 # ^ POH ^ 무신 #이다



AEDA

!@ \* PER \* 이순신 @ 은 조선시대 : 중기 # ^ POH ^ 무신 # ?이다

# Data Augmentation – Entity Swap

토론 게시판 강석민 캠퍼님의 포스트를 기반으로 데이터 생성 : <https://stages.ai/competitions/75/discussion/talk/post/657>

sbj obj를 바꿔도 문제없는 라벨	no_relation   org:alternate_names   per:alternate_name   per:other_family   per:colleagues   per:siblings   per:spouse
sbj, obj를 바꾸면 라벨을 바꿔야 하는 경우	org:members ↔ org:members_of per:children ↔ per:parents
sbj, obj 한쪽 방향만 바꿔야 하는 경우	org:top_members/employees → per:employee_of

< ex) no\_relation label >

<Something>는 #^PER^조지 해리슨#이 쓰고 @\*ORG\*비틀즈@가 1969년 앨범 《Abbey Road》에 담은 노래다.



<Something>는 @\*PER^조지 해리슨@이 쓰고 #^ORG^비틀즈#가 1969년 앨범 《Abbey Road》에 담은 노래다.

# Data Augmentation – Random Delete

---

[CLS] @ \* PER \* 이순신 @ 과 # ^ POH ^ 무신 # 의 관계 [SEP] @ \* PER \* 이순신 @ 은 조선 시대 중기  
# ^ POH ^ 무신 #이다 [SEP]



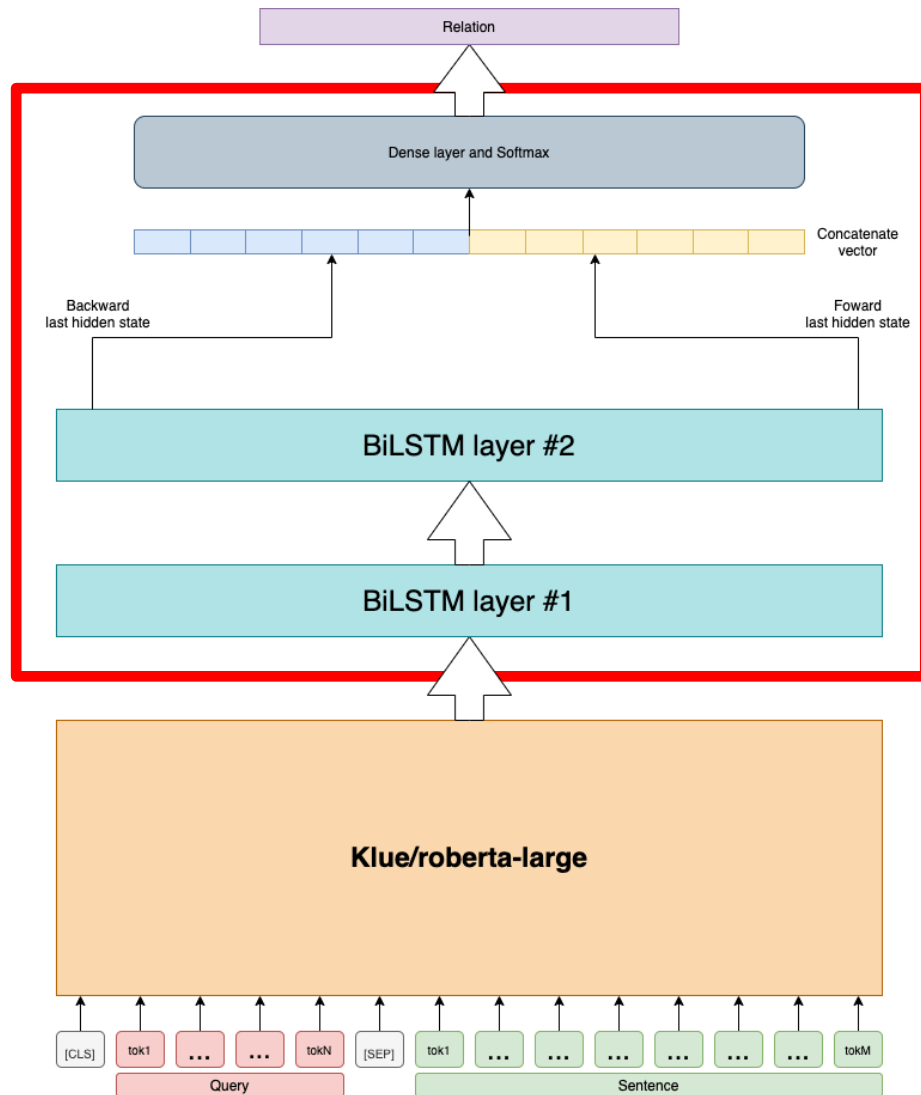
Random Delete

[CLS] @ \* PER \* 이순신 @ 과 # ^ POH ^ 무신 # 의 관계 [SEP] @ \* PER \* 이순신 @ 은 조선 중기  
# ^ POH ^ 무신 #이다 [SEP]

30% 확률로 Delete



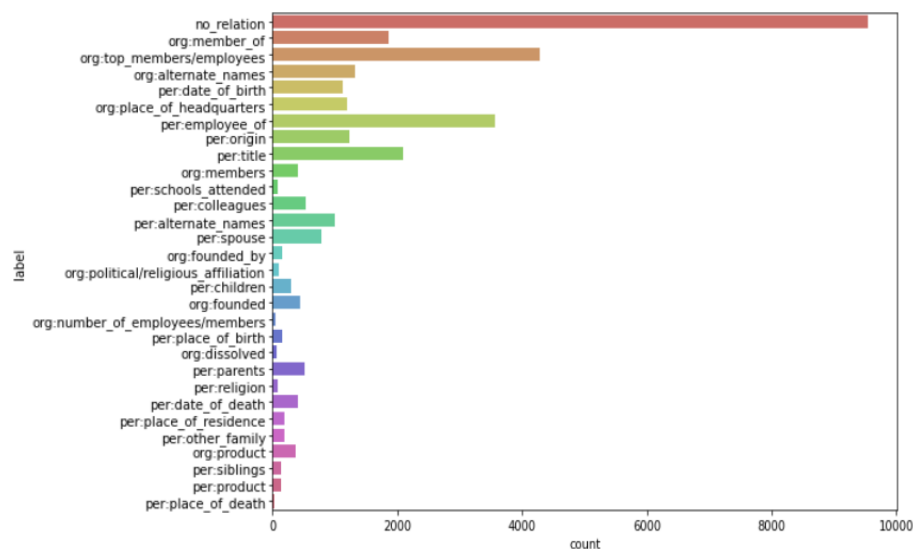
# Model Architecture



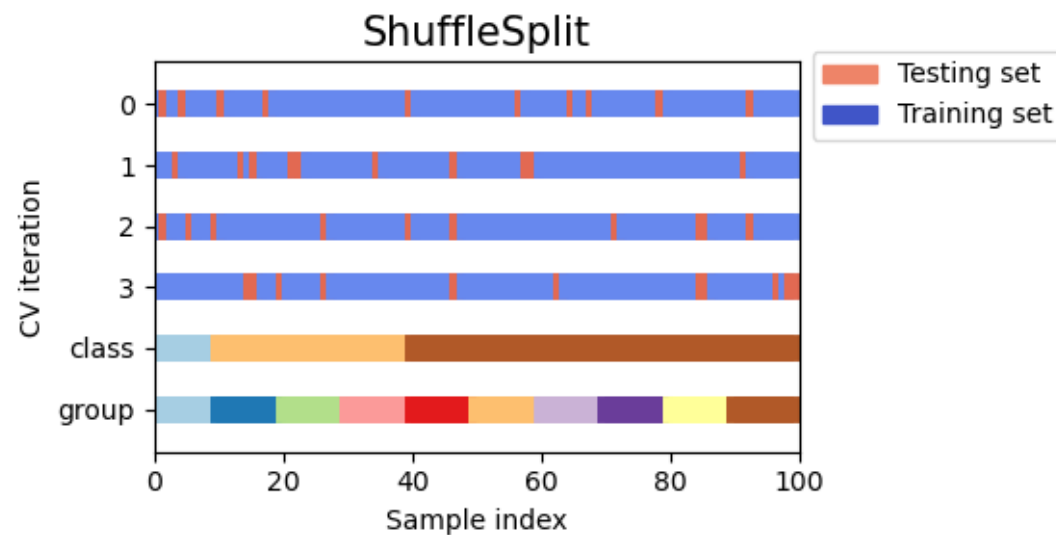
- Loss function : Label Smoothing ( $\epsilon=0.1$ )
- Optimizer : AdamW (Weight decay=0.01)
- Learning rate Scheduler : Warmup and linear decay ( $1e-5$ )

# Ensemble

## Stratified K-Fold & OOF(Out-of-Fold) Prediction



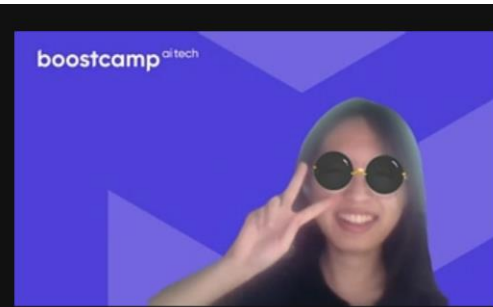
레이블 분포가 불균형 함을 확인



원본 데이터와 유사한 분포를 train, val set에도 유지하기 위해 **Stratified K-Fold** 채택

# 쿼터백's member

# 민지



# 석민



# 재형



# 성호



# 동건



# 다인



# 다영



# 감사합니다

쿼터백 조

김다영 \* 김다인 \* 박성호 \* 박재형 \* 서동건 \* 정민지 \* 최석민

Q n A