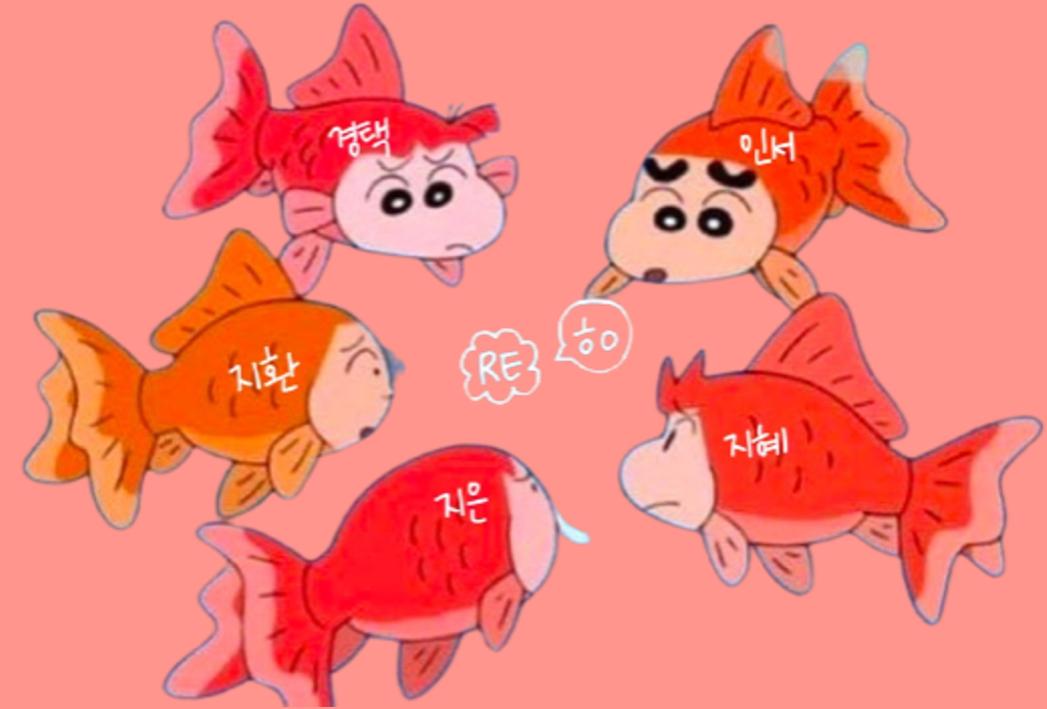


문장 내 개체간 관계 추출 대회

NLP-12- 훈제연어들

송인서, 윤지환, 박지은, 박경택, 문지혜



INDEX

1. 개요

1.1 대회 소개

1.2 팀 소개

2. 과정

2.1 프로젝트 flow chart

2.2 Data EDA

2.3 Settings

2.4 Experiments

2.5 협업

3. 결과 및 고찰

3.1 잘한 점

3.2 아쉬운 점

3.3 질문

1. 개요

1.1 대회 소개

1.2 팀 소개

1.1 대회 소개

관계 추출(Relation Extraction)

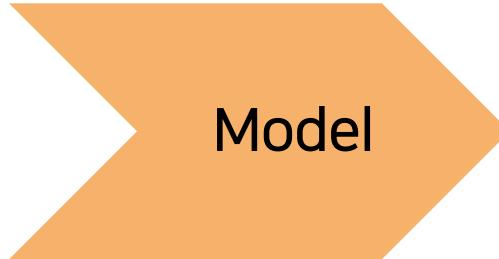
: 문장 내 단어(Entity)에 대한 속성과 관계를 예측하는 Task

Input

sentence : 뉴질랜드 1차산업부 생물보안 대변인 캐서린 더시 박사는 과일파리를 잡기 위해 오클랜드 전역에 1천300여개의 파리통을 놓았으며, 곧 통들을 마지막으로 점검할 것이라고 말했다.

subject_entity : 뉴질랜드

object_entity : 오클랜드



Output

pred_label : org:members

probs : "[0.21730691194534302, ..., 0.00018276390619575977]"

1.2 팀 소개



문지혜

- EDA
- loss function 구현 및 실험
- PLM 모델 성능 비교



박경택

- 프로젝트 전반부 PM
- Custom model 구현 & 실험
- 모델 구조 변화 실험



박지은

- main code refactoring
- prompt 구현 및 실험
- custom model 실험 및 사후 분석



윤지한

- entity marker 구현
- TAPT, entity embedding, lr scheduler 실험
- 모델학습 및 양상블



송인서

- 프로젝트 후반부 PM
- main code refactoring
- sweep 구현

2. 과정

2.1 프로젝트 flow chart

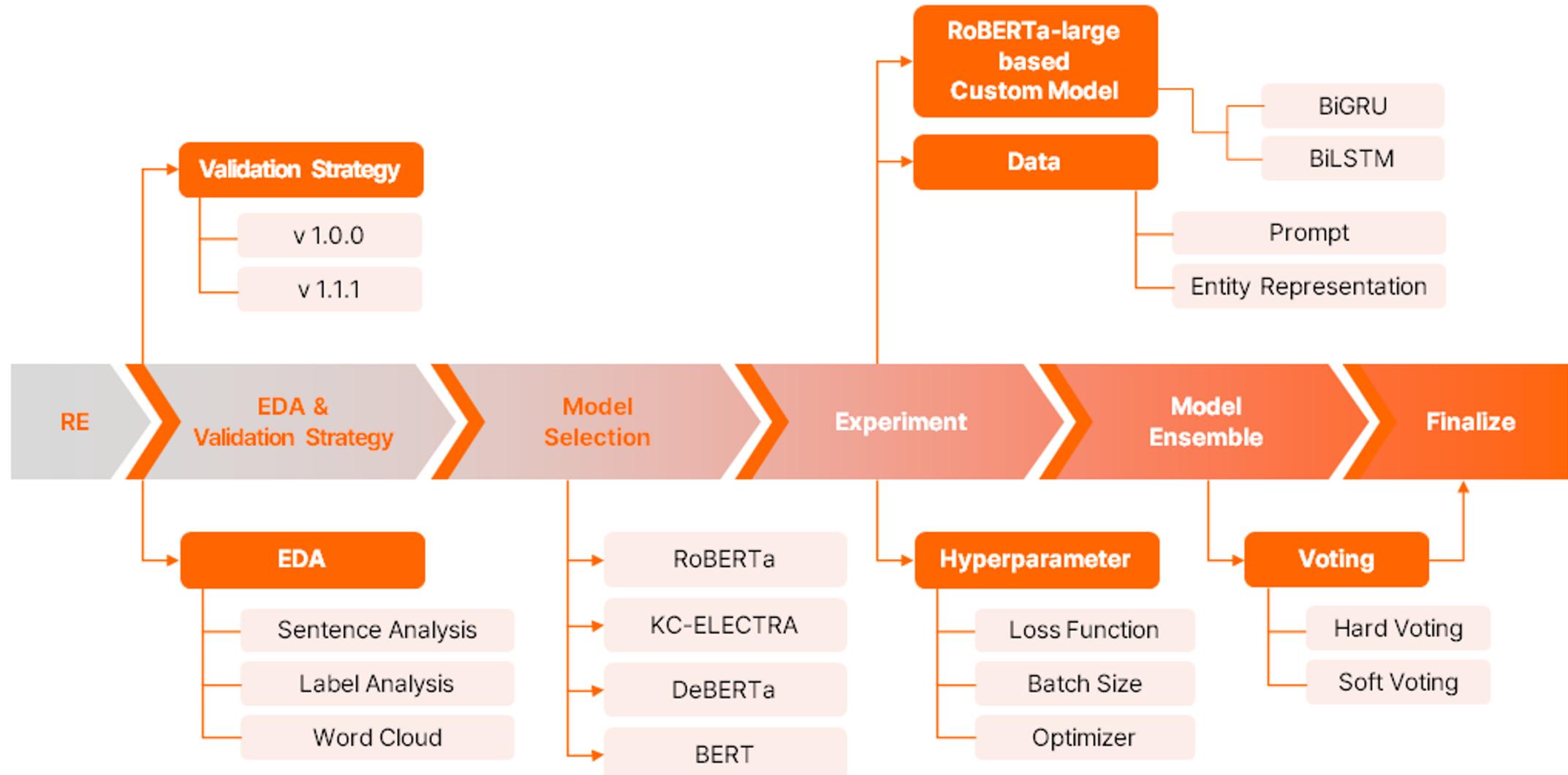
2.2 Data EDA

2.3 Settings

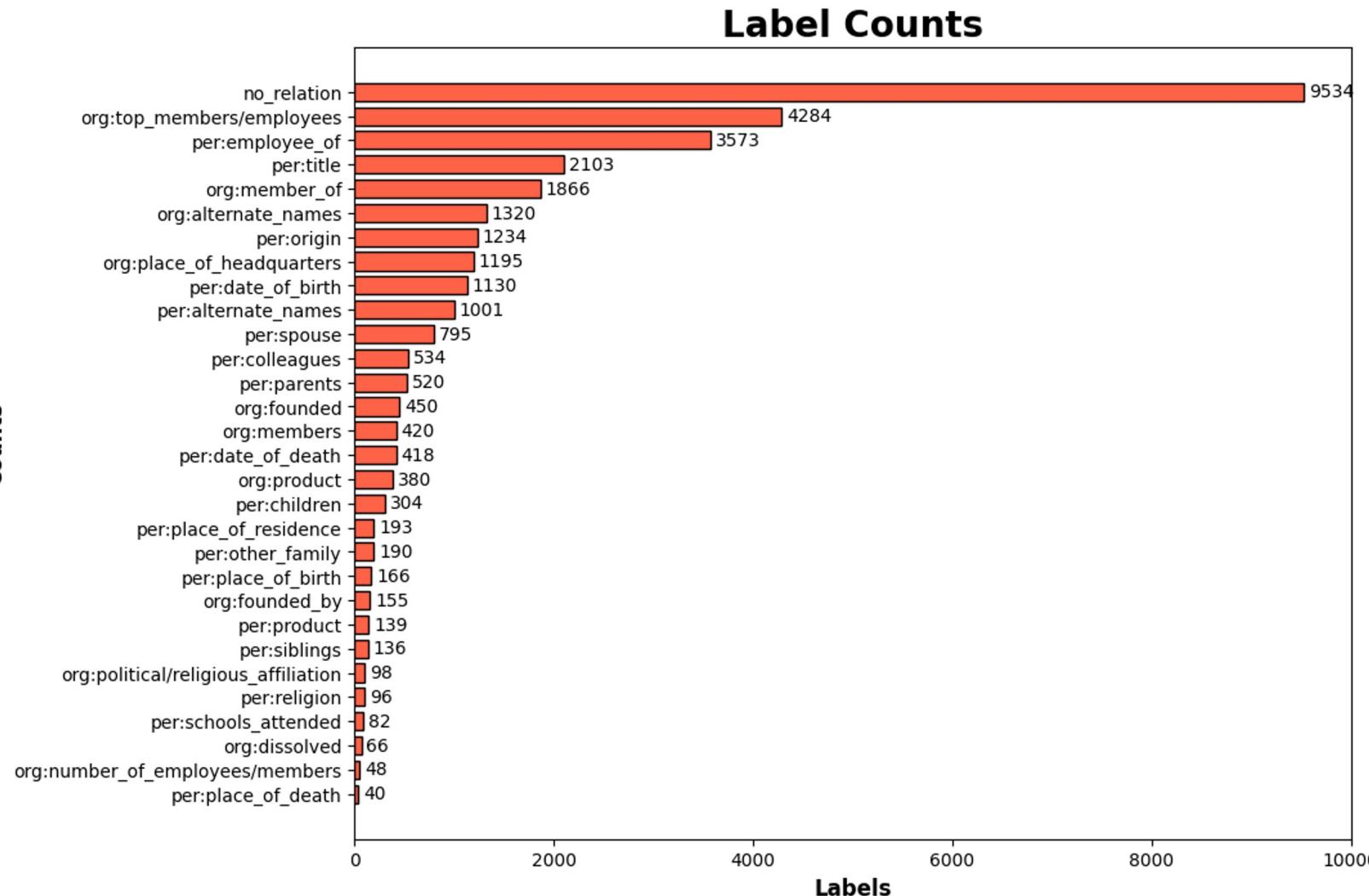
2.4 Experiments

2.5 협업

2.1 프로젝트 Flow Chart



2.2 Data EDA



Class imbalance 발견

no_relation label : 9,534개

org:top_members : 4,284개

per:employee_of : 3,573개

2.3 Settings

- Validation Set: test set에 대한 inference 분포와 유사하도록 설정
- Base Model: klue/RoBERTa-large model

KLUE: Korean Language Understanding Evaluation

Table 31: Implementation details of KLUE-BERT and KLUE-RoBERTa. WWM refers to the whole word masking strategy.

Model	# Parameter	Masking	Training Steps	Batch Size	Learning Rate	Device
KLUE-BERT _{BASE}	110M	Static, WWM	1M	256	10 ⁻⁴	TPU v3-8
KLUE-RoBERTa _{SMALL}	68M	Dynamic, WWM	1M	2048	10 ⁻⁴	8× V100 GPUs
KLUE-RoBERTa _{BASE}	110M	Dynamic, WWM	1M	2048	10 ⁻⁴	8× V100 GPUs
KLUE-RoBERTa _{LARGE}	337M	Dynamic, WWM	500k	2048	10 ⁻⁴	8× V100 GPUs

Model	Explanation	Micro F1 Score	AUPRC
klue/roberta-large	KLUE Dataset으로 RoBERTa 모델을 사전학습한 모델	71.371	69.653
klue/bert-base	KLUE Dataset으로 BERT 모델을 사전학습한 모델	66.751	64.76
beomi/KcELECTRA-base-v2022	네이버 뉴스에서 댓글과 대댓글을 수집해, 토크나이저와 ELECTRA 모델을 처음부터 학습한 Pretrained ELECTRA 모델	61.722	46.494
lighthouse/mdeberta-v3-base-kor-further	microsoft 가 발표한 mDeBERTa-v3-base 모델을 약 40GB의 한국어 데이터에 대해서 추가적인 사전학습을 진행한 언어 모델	65.571	55.646

2.4 Experiments

[Data]

- Prompt
- Entity Representation

[Custom Model]

- RoBERTa + BiGRU
- RoBERTa + BiLSTM

[Hyperparameter]

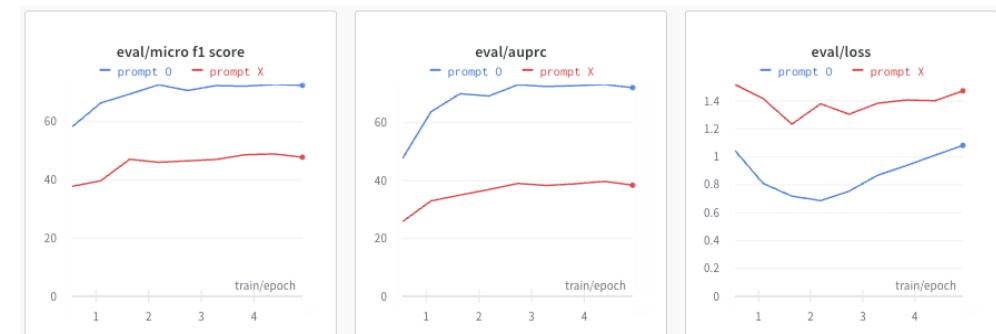
- Loss Function
- Batch Size

2.4 Experiments - Data part: Prompt



```
# load_data.py
def tokenized_dataset(dataset, tokenizer):
    """ tokenizer에 따라 sentence를 tokenizing 합니다."""
    concat_entity = []
    for e01, e02 in zip(dataset['subject_entity'], dataset['object_entity']):
        temp = ''
        temp = e01 + '[SEP]' + e02
        concat_entity.append(temp)
    tokenized_sentences = tokenizer(
        concat_entity,
        list(dataset['sentence']),
        return_tensors="pt",
        padding=True,
        truncation=True,
        max_length=256,
        add_special_tokens=True,
    )
    return tokenized_sentences
```

prompt	micro f1	auprc
X	47.822	38.445
O	72.193 (24.3 ▲)	72.167 (33.8 ▲)



2.4 Experiments - Data part: Prompt

Relation Extraction Based on Prompt Information and Feature Reuse

Ping Feng^{1,2,3*}, Xin Zhang², Jian Zhao^{2,3}, Yingying Wang², Biao Huang²

¹ Jilin University, Changchun Jilin 130012, China

² Changchun University, Changchun Jilin 130022, China

³ Jilin Provincial Key Laboratory of Human Health State Identification and Function Enhancement, Changchun Jilin 130022, China

Citation: Feng, P., Zhang, X., Zhao, J. et al.: Relation extraction based on prompt information and feature reuse. Data Intelligence. 5 (2023). DOI: https://doi.org/10.1162/dint_a_00192

Method	Input Example	klue/roberta-large	
		Micro F1 Score	AUPRC
default	[CLS] 부덕이는 부캠의 마스코트이다. [SEP]	47.822	38.445
s_sep_o	[CLS] 부덕이 [SEP] 마스코트 [SEP] 부덕이는 부캠의 마스코트이다. [SEP]	72.193	72.167
s_and_o	[CLS] 부덕이와 마스코트의 관계 [SEP] 부덕이는 부캠의 마스코트이다. [SEP]	73.575	71.588
quiz	[CLS] 다음 문장에서 부덕이와 마스코트의 관계를 추출하시오. [SEP] 부덕이는 부캠의 마스코트이다. [SEP]	73.299	71.995
question	[CLS] 다음 문장에서 부덕이와 마스코트의 관계를 추출하시오. [SEP] 부덕이는 부캠의 마스코트이다. 부덕이와 마스코트는 어떤 관계입니까? [SEP]	73.446	73.399

2.4 Experiments - Data part: Entity Representation

An Improved Baseline for Sentence-level Relation Extraction

Wenxuan Zhou

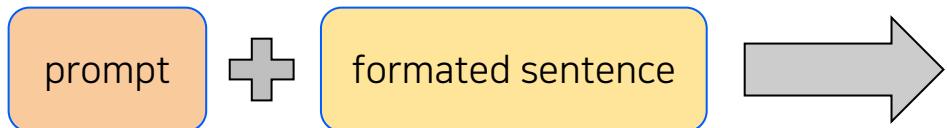
University of Southern California
zhouwenx@usc.edu

Muhao Chen

University of Southern California
muhaoche@usc.edu

Method	Input Example	klue/roberta-large	
		Micro F1 Score	AUPRC
Default	부덕이는 부캠의 마스코트이다.	47.822	38.445
Entity Mask	[SUBJ-PER]는 부캠의 [OBJ-TITLE]이다.	68.057	66.13
Entity Marker	[E1] 부덕이 [/E1]는 부캠의 [E2] 마스코트 [/E2]이다.	72.678	72.951
Entity Marker (punct)	@ 부덕이 @ 는 부캠의 # 마스코트 # 이다.	72.458	72.548
Typed entity marker	<S:PERSON> 부덕이 </S:PERSON>는 부캠의 <O:TITLE> 마스코트</O:TITLE>이다.	73.005	72.647
Typed entity marker (punct)	@ * person * 부덕이 @ 는 부캠의 # ^ title ^ 마스코트 # 이다.	72.54	72.14

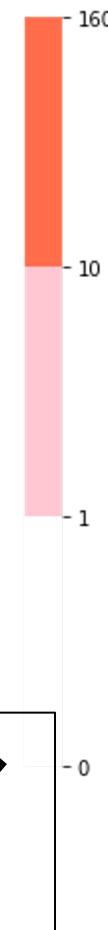
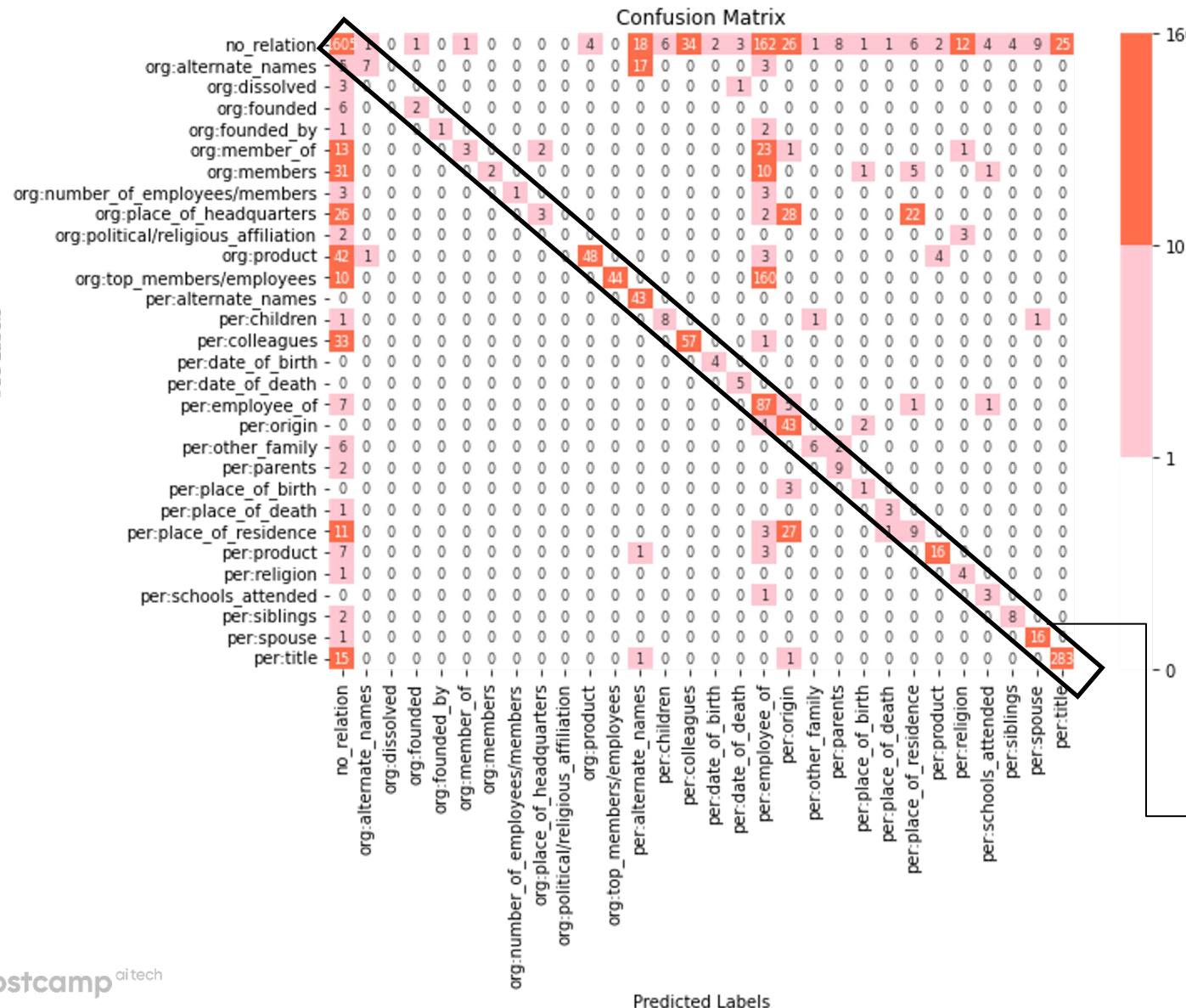
2.4 Experiments - Data part: Prompt + Entity Representation



- prompt를 통해 모델이 진행할 task를 직접적으로 명시해줌으로써 성능이 크게 개선되었다.
- Entity representation을 통해 문장 내에서 entity를 강조해줌으로써 성능이 크게 개선되었다.
- 여러 방법 중 특정한 하나를 고르기엔 성능 차가 크게 유의미하지 않았다.

	input_format	prompt	micro f1	auprc	loss
1	default	default	47.822	38.445	1.473
2	default	s_sep_o	72.193	72.167	1.021
3	default	s_and_o	73.575	71.588	1.032
4	default	quiz	73.299	71.995	1.023
5	default	problem	73.446	73.399	1.039
6	entity_mask	default	68.057	66.13	1.057
7	entity_marker	default	72.678	72.951	0.9561
8	typed_entity_marker	default	73.005	72.647	0.9917
9	typed_entity_marker_punct	default	72.54	72.14	1.082
10	typed_entity_marker_punct_한글	default			
11	entity_marker	s_sep_o	72.356	71.557	1.056
12	entity_marker	s_and_o	72.984	71.359	1.022
13	typed_entity_marker	s_sep_o	72.968	71.612	1.061
14	typed_entity_marker	s_and_o	73.777	71.62	1.021
15	typed_entity_marker_punct	s_sep_o	72.074	73.906	1.129
16	typed_entity_marker_punct	s_and_o	72.94	71.698	1.082

2.4 Experiments - Data part



Model: klue/roberta-large

Loss function: CrossEntropyLoss

Prompt: s_and_o

Entity representation: typed_entity_marker_punct

Predicted Labels = True Labels
올바르게 예측한 수

2.4 Experiments - Data part



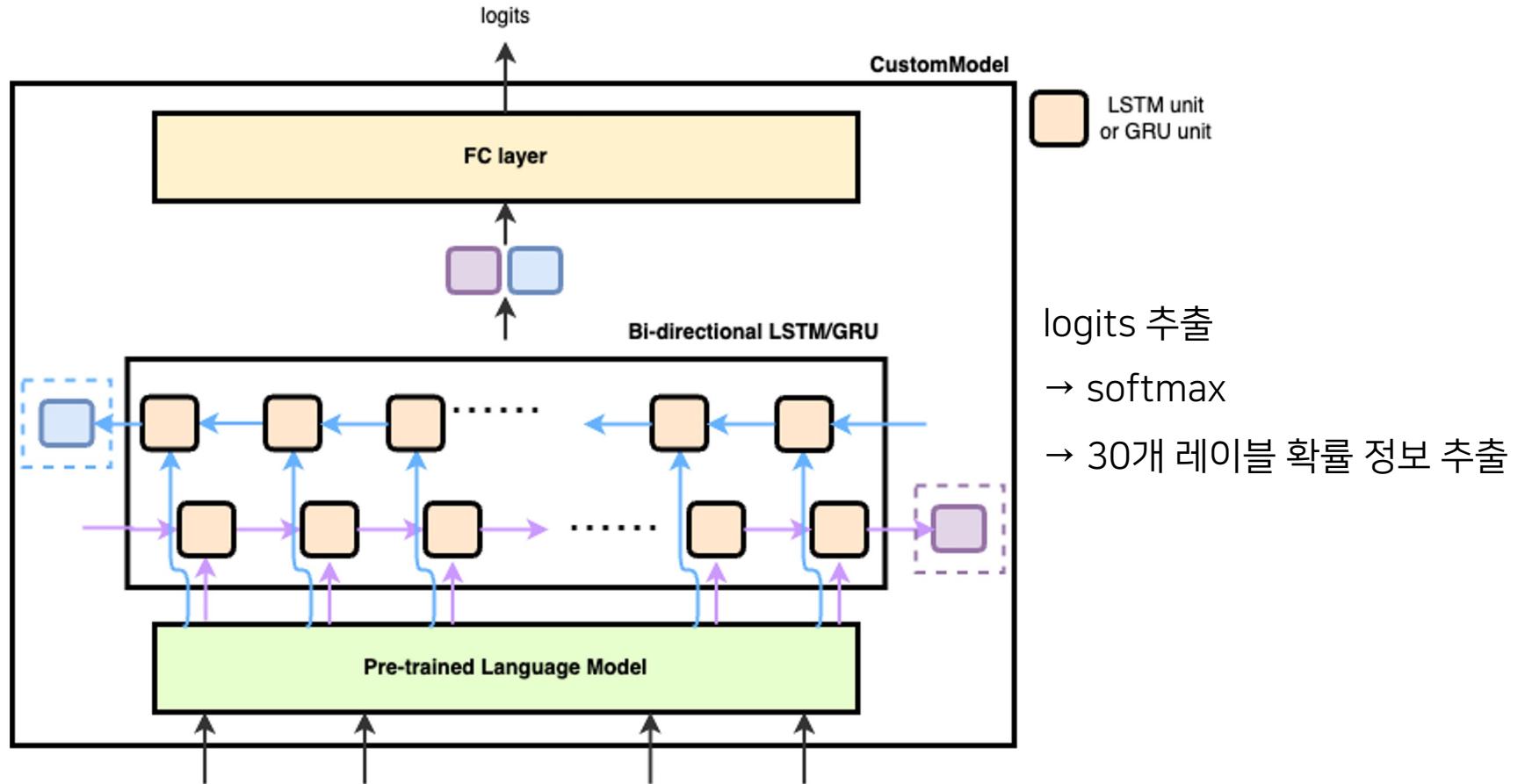
#	True Labels	Predicted Labels	value
1	per:place_of_residence	per:origin	27
2	org:place_of_headquarters	per:origin	28
3	org:place_of_headquarters	per:place_of_residence	22
4	org:alternate_names	per:alternative_names	17
5	org:member_of	per:employee_of	23
6	org:members	per:employee_of	10
7	org:top_members/employees	per:employee_of	160

- 1 : 유사한 label
- 7 : 포함관계의 label
- 2, 3, 4, 6 : entity type (per / org) 간의 차이만 존재하는 label

모델이 데이터의 전반적인 특징은 파악하고 있으나, 세부적인 부분까지 엄밀하게 판단하지는 못함

2.4 Experiments - Model part

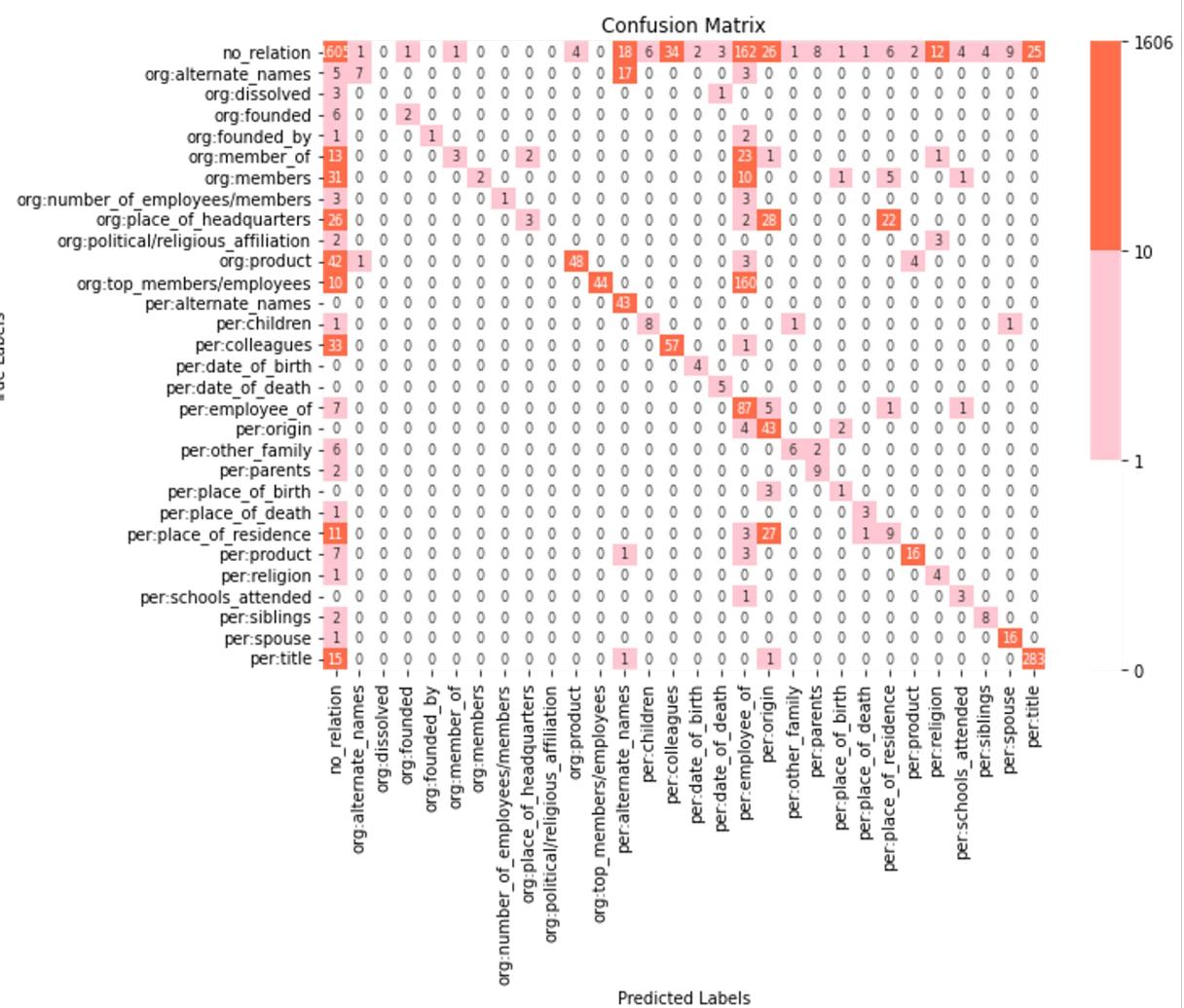
✓ RoBERTa-large Based Custom Model



[CLS], 부, ##덕, ##이, ##와 마스코트, ##의, 관계, [SEP], 부, ##덕, ##이, ##는, 부, ##캠, ##의, 마스코트, ##이다, ., [SEP]

2.4 Experiments - Model part

기존 모델 (RoBERTa-Large)



실험 모델 (RoBERTa-Large + BiGRU)

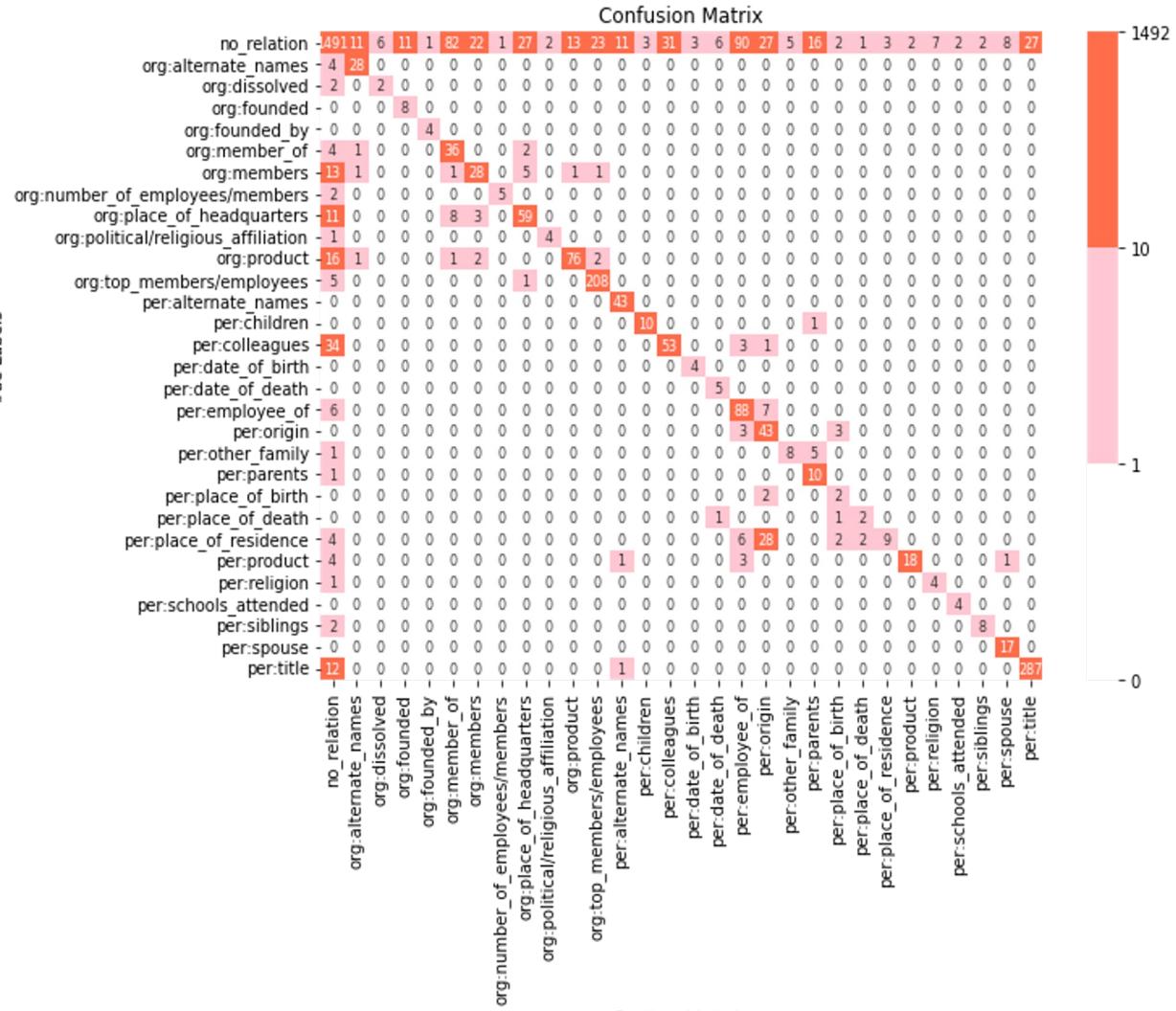


2.4 Experiments - Model part

기존 모델 (RoBERTa-Large)

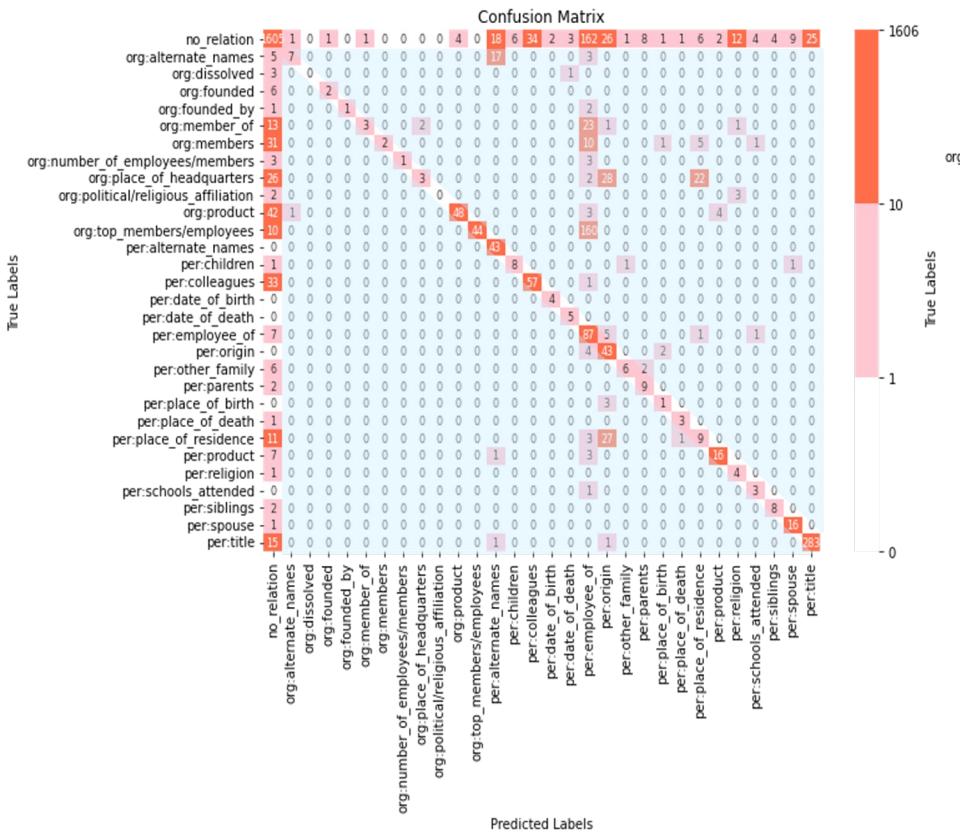


실험 모델 (RoBERTa-Large + BiLSTM)

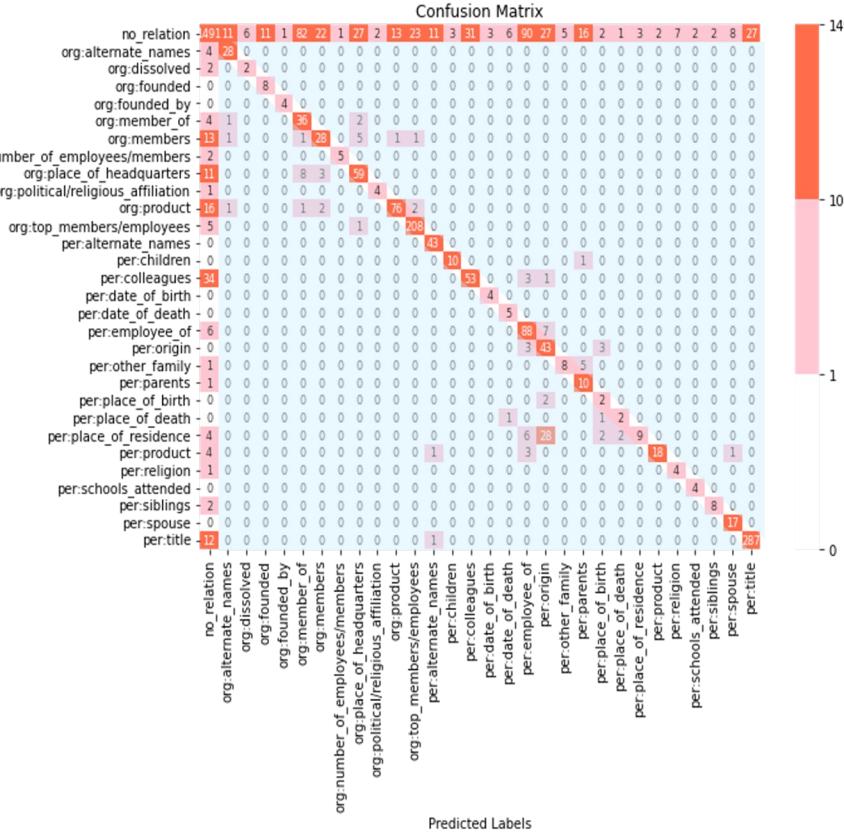


2.4 Experiments - Model part

기존 모델 (RoBERTa-Large)



실험 모델 (RoBERTa-Large + BiLSTM)

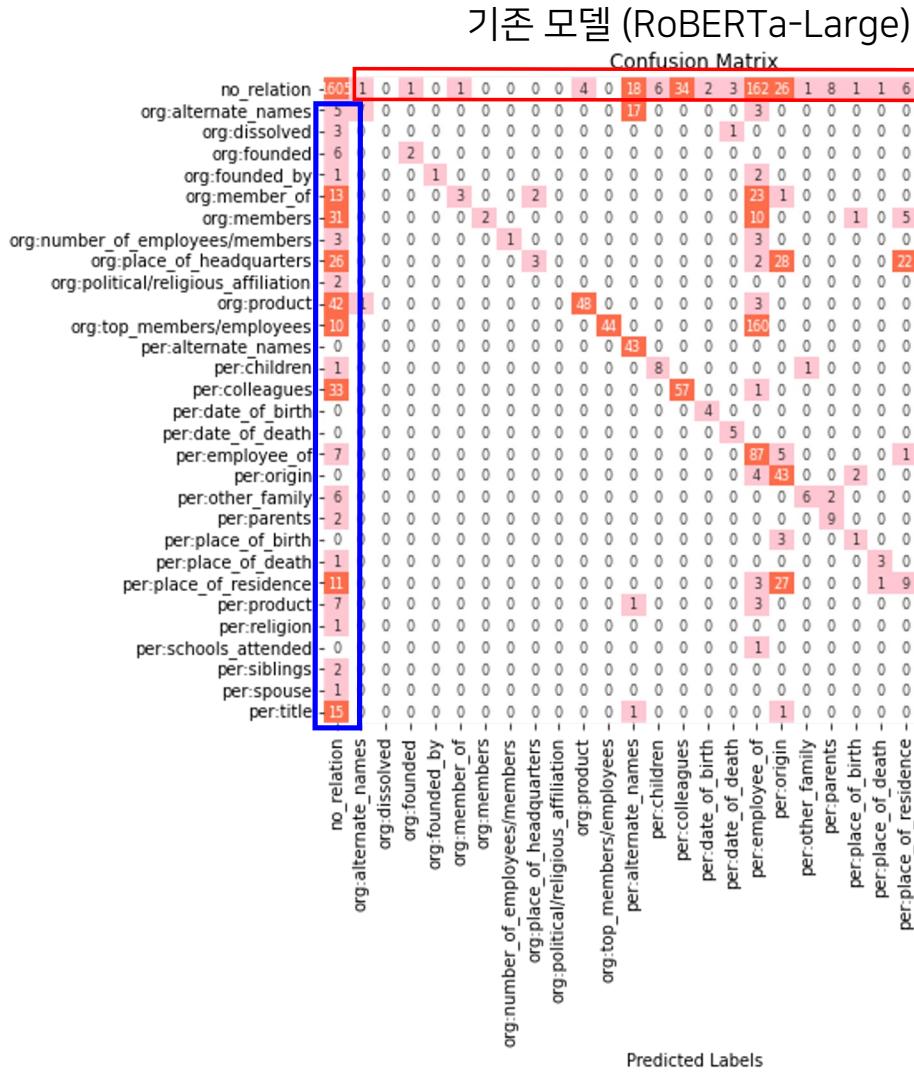


	기존 모델	BiGRU	BiLSTM
FP	683	583	546
FN	581	269	224
precision	0.5118	0.6381	0.6628
recall	0.5520	0.7926	0.8273
f1 score	0.5312	0.7070	0.7359

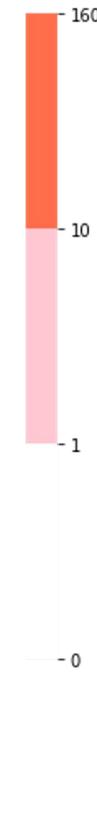
False Positive, False Negative 감소
-> precision, recall, f1 score 증가

2.4 Experiments - Model part

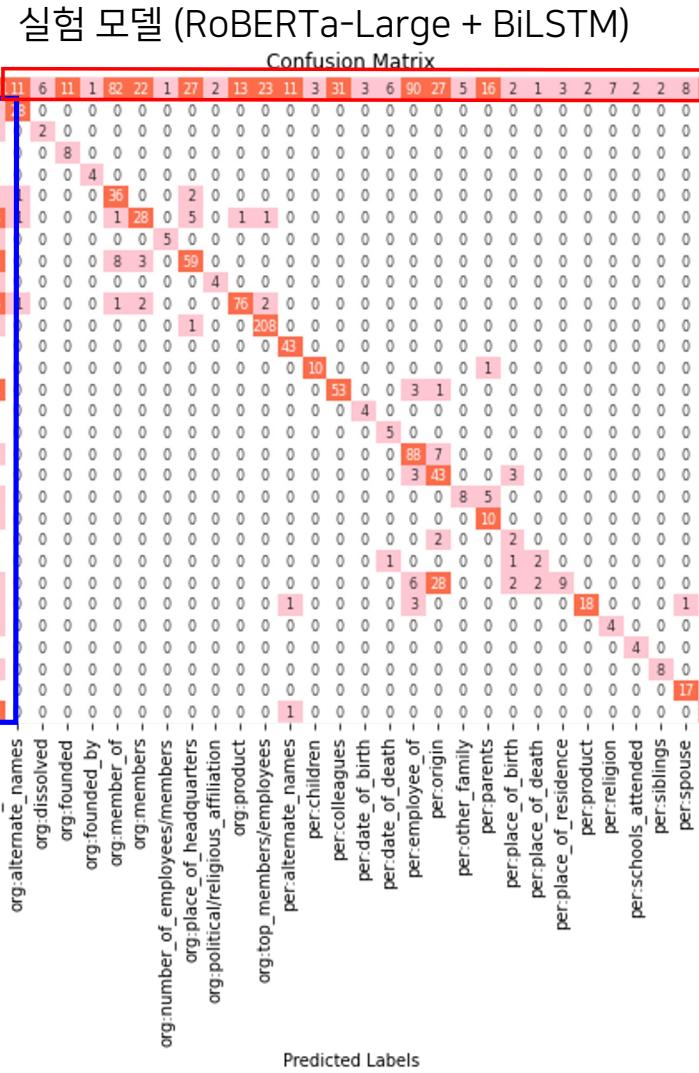
True Labels



정답값은 no_relation이지만
다른 label로 예측한 값들 (FN)

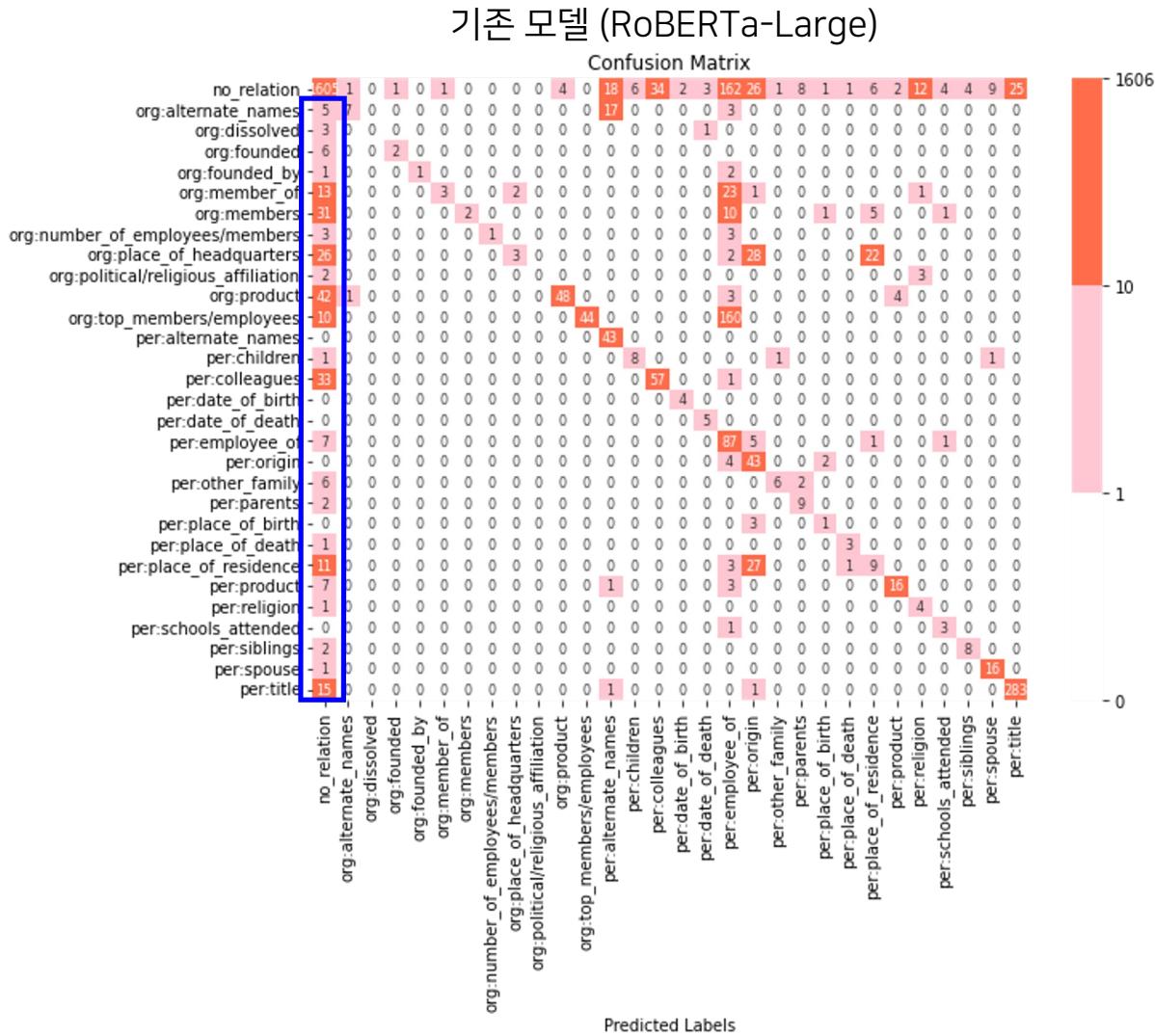


정답값은 no_relation이 아닌데
no_relation으로 예측한 값들 (FP)



2.4 Experiments - Model part

True Labels



		True Class	
		no_relation	no_relation x
Predicted Class	no_relation	TP	FP
	no_relation x	FN	정답값은 no_relation인데 다른 label로 예측한 값
			TN

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{up arrow}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{down arrow}$$

=> 낮은 precision, 높은 recall

2.4 Experiments - Model part

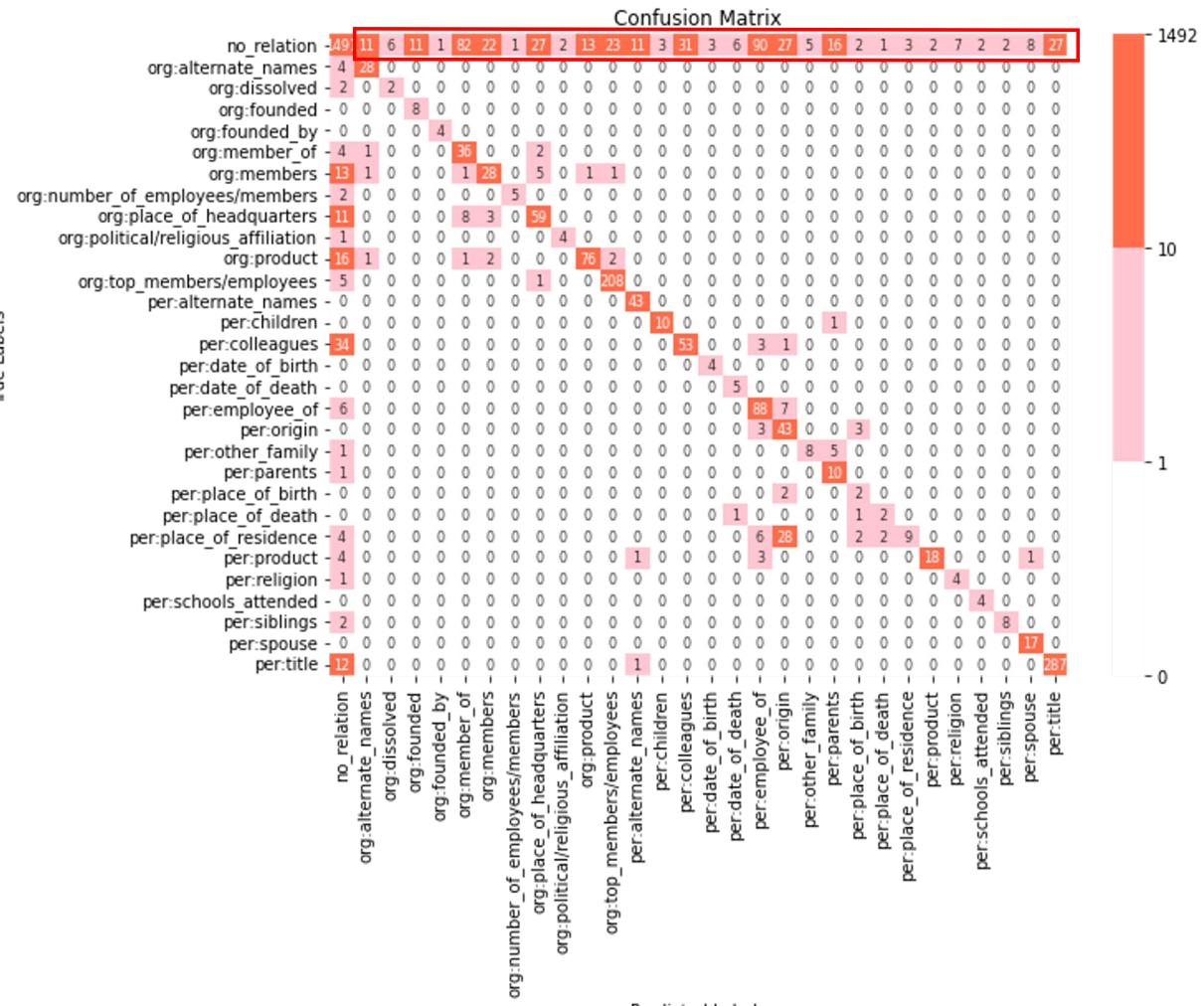
		True Class	
		no_relation	no_relation x
Predicted Class	no_relation	TP	FP 정답값은 no_relation이 아닌데 no_relation으로 예측한 값
	no_relation x	FN 정답값은 no_relation인데 다른 label로 예측한 값	TN

$$Precision = \frac{TP}{TP + FP} \downarrow$$

$$Recall = \frac{TP}{TP + FN} \uparrow$$

=> 높은 precision, 낮은 recall

실험 모델 (RoBERTa-Large + BiLSTM)



2.4 Experiments - Model part

- no_relation label에 대한 recall 및 precision

	기존 모델	Bi-GRU	Bi-LSTM
precision	0.8751	0.9020	0.9238
recall	0.8290	0.7557	0.7701

- 기존 모델 < Bi-GRU, Bi-LSTM
- 기존 모델 > Bi- GRU, Bi-LSTM

- Leaderboard 제출 결과

	기존 모델	Bi-GRU	Bi-LSTM
Micro f1	74.3558	75.2809	76.5345
auprc	76.6053	80.2392	79.9067

2.4 Experiments - Model part

		True Class	
		no_relation	no_relation x
Predicted Class	no_relation	TP 정답값은 no_relation이 아닌데 No_relation으로 예측한 값	FP 정답값은 no_relation이 아닌데 No_relation으로 예측한 값
	no_relation x	FN 정답값은 no_relation인데 다른 label로 예측한 값	TN

Micro F1 score

- micro-precision과 micro-recall의 조화 평균이며, 각 샘플에 동일한 importance를 부여해, 샘플이 많은 클래스에 더 많은 가중치를 부여합니다. 데이터 분포상 많은 부분을 차지하고 있는 no_relation class는 제외하고 F1 score가 계산 됩니다.

$$precision = \frac{TP}{TP + FP}$$

- 다른 class의 precision, recall 값에 영향을 미치므로, micro f1 score에 영향을 미침

$$Recall = \frac{TP}{TP + FN}$$

- 다른 class의 precision, recall에 영향을 미치지 못하므로, micro f1 score에도 영향을 미치지 못함

2.4 Experiments - Hyperparameter part

- Loss Function

EDA 과정에서 데이터에 class imbalance가 있음을 발견

→ 기존에 사용하던 cross-entropy loss 대신, class imbalance 데이터를 훈련하는 데 효과적인 것으로 알려진 focal loss를 사용해보자!

Cross-entropy Loss		Focal Loss
$\text{CE}(p_t) = -\log(p_t)$.	수식	$\text{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$.
71.371	Micro F1 Score	71.225
69.653	AUPRC	71.500

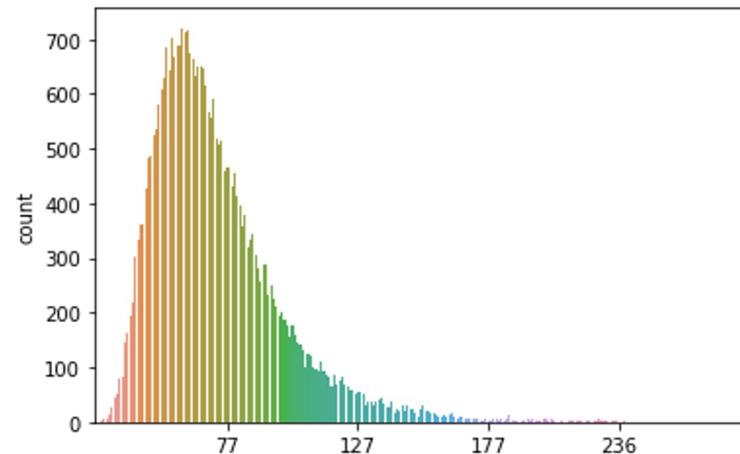
지난 5일 아이즈원 해외 팬은 자신이 강혜원 씨 이름으로 스코틀랜드에 있는 땅을 샀다는 글과 함께 인증 사진을 게시했다.

{'word': '강혜원', 'start_idx': 21, 'end_idx': 23, 'type': 'PER'}

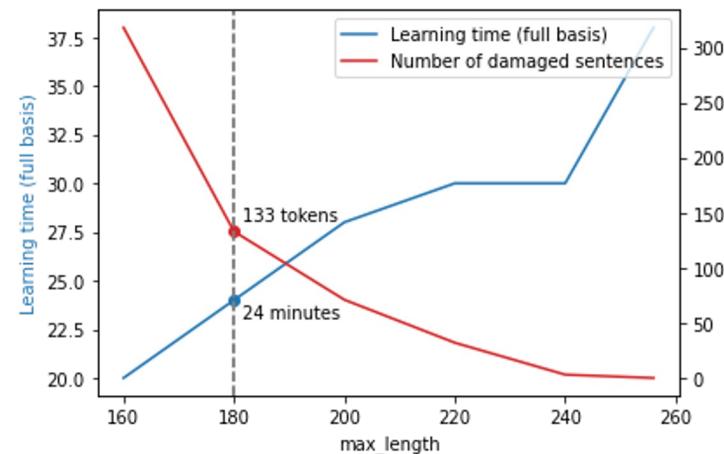
{'word': '아이즈원', 'start_idx': 6, 'end_idx': 9, 'type': 'POH'} per:employee_of wikitree

2.4 Experiments - Hyperparameter part

- Batch size



- special token을 제외한 input_id의 길이는 180 이하에 대부분 몰려있음



- tokenizer의 max_length 변화에 따른 학습시간과 손상되는 sentence 수

	적용 전	적용 후
학습시간 (min)	34	24
micro f1	74.241	75.226
auprc	76.5932	77.5233

- tokenizer의 max_length 변화에 따른 학습시간과 손상되는 sentence 수

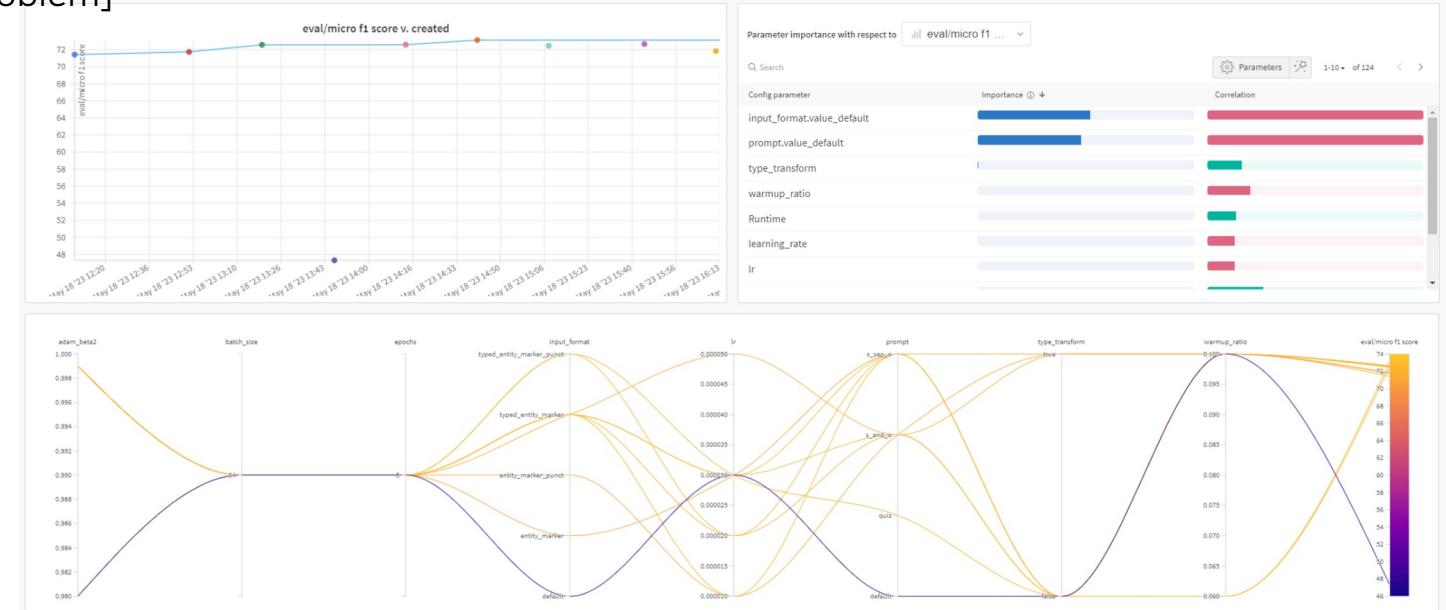
학습시간과 손실되는 전체 token 간의 trade-off를 고려해
max_length를 180으로 줄이고 batch size를 64로 늘림

후에 gradient accumulation도 시도하였으나 큰 차이는 없었음

2.4 Experiments - 최종 모델

- Parameter Tuning and Voting

- WandB Sweep 활용한 Bayesian search
- Base, Bi-LSTM, Bi-GRU 각각 tuning 진행
- 상위 5개 모델 hard voting
- Config.yaml
 - input_format : [default, entity_marker, entity_marker_punct, typed_entity_marker, typed_entity_marker_punct]
 - prompt : [default, s_sep_o, s_and_o, quiz, problem]
 - type_transform : [true, false]
 - lr : [1e-5, 2e-5, 3e-5, 5e-5]
 - epochs : [3, 5]
 - batch size : 64
 - adam_beta2 : [0.98, 0.999]
 - warmup_ratio : [0.06, 0.1]



2.5 협업



Hugging Face

- Hugging Face Datasets을 통한 데이터셋 버전 컨트롤

The screenshot shows the Hugging Face Datasets interface for the 'RE_Competition' dataset. The top navigation bar includes 'Datasets', 'Smoked-Salmon-s/RE_Competition' (private), 'Dataset card', 'Files' (selected), 'Community', and 'Settings'. Below the navigation is a dropdown for branches ('main') and the dataset name ('RE_Competition'). It shows 1 contributor and 4 commits. A button for '+ Add file' is visible. The main area displays four CSV files: '.gitattributes', 'test.csv', 'train.csv', and 'valid.csv', each with its size, commit hash, version, and last modified time.

File	Size	Commit Hash	Version	Last Modified
.gitattributes	2.31 kB	a6a27f7	v1.0.0	15 days ago
test.csv	3 MB		v1.0.0	15 days ago
train.csv	12 MB (LFS)		v1.1.1	9 days ago
valid.csv	1.36 MB		v1.1.1	9 days ago

2.5 협업

A screenshot of a GitHub Issues page with a dark theme. The header shows 4 Open and 14 Closed issues. The list includes several closed pull requests with titles like "[FEAT] loss function에 class별 가중치 적용하기", "[FEAT] prompt 고도화", "[FEAT] Roberta TAPT", "[FEAT] initialization & regularization 실험 등", "[FIX] Batch size 64로 올리기", "[FEAT] BiGRU가 부가된 custom model 구현", and "[FIX] entity type 한글로 변경하는 기능 추가". Each item has a small icon, a timestamp, and a link to the pull request.

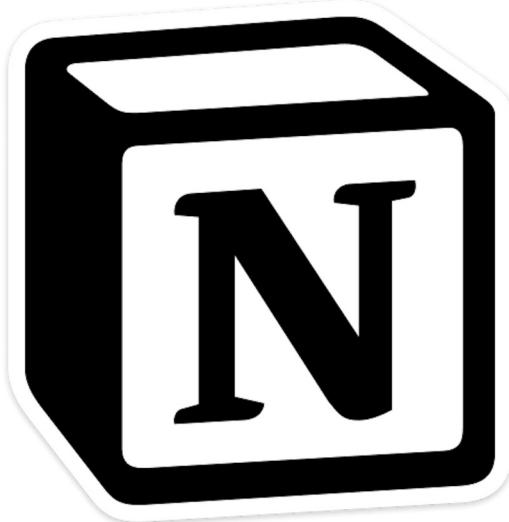
A screenshot of a GitHub Issues page with a dark theme. The header shows 0 Open and 11 Closed issues. The list includes merged pull requests with titles like "feat/#22 additional prompt", "Feat/#23 weights loss", "Feat/early stopping", "Feat/#8 initialization, regularization, normalization", "Feat/#8 bilstm model", and "Feat/#13 bigru model". Each item has a small icon, a timestamp, and a link to the pull request.



GitHub flow

- GitHub flow
- Convention
 - Commit message
 - Branch name
 - Issue template
 - PR template
- Issue 및 pull request 문화 정착

2.5 협업



- bulletin board
- 실험 기록

The screenshot shows a digital bulletin board with three main sections: 'Not started' (2 items), 'In progress' (16 items), and 'Done' (33 items). Each item has a title, a list of contributors (with small profile icons), and a brief description.

- Not started:** 0518 회의록 + 멘토링 내용, R-drop
- In progress:** 강의 들어주세요, EDA - base, Validation Strategy, 사후 분석 - base, Bi-GRU RE Model, Bi-LSTM RE Model
- Done:** KLUE 대회 안내, AI Stages 대회 소개, 0502 회의록, AI Stages 개시판, 0503 회의록, baseline code refactoring, EDA - label analysis, 0504 회의록

Date	Status	Tags	Category	Aa Name	Person	Public Micro F1	Public AUPRC
May 8, 2023	Done	Experiments	Data	dev set test 2 (train:dev=8:2)	박경택	68.2943	74.1954
May 8, 2023	Done	Experiments	Model	LB 73.3458 baseline setti	지환 윤	74.7337	73.3458
May 8, 2023	Done	Experiments	Data	dev set test 1 (failed)	박경택	65.4252	65.1374
May 8, 2023 →	Done	Experiments	Model	Loss function & Custom Trainer	Jihye		
May 9, 2023	Done	Experiments	Model	Custom LR_scheduler	지환 윤		
May 9, 2023	Done	Experiments	Data	dev set test (v1.0.0)	박지은	68.2736	76.3754
May 10, 2023	Done	Experiments	Data	dev set test (v1.1.1)	박지은	68.0290	77.3309
May 10, 2023	Done	Experiments	Model	focal loss alpha=0.25 & gamma=2	박경택	71.3011	75.9856
May 11, 2023	Done	Experiments	Data	Entity Representation & prompt	박지은, 지환 윤		
May 13, 2023	Done	Experiments	Model	batch size 늘리보자	지환 윤	75.2226	77.5233
May 15, 2023	Done	Experiments	Model	loss.py 정리	Jihye		
May 15, 2023	In progress	Experiments	Model	Bi-LSTM RE Model	박지은	76.5345	79.9067
May 16, 2023	Done	Experiments	Model	weighted loss	Jihye		
May 17, 2023	In progress	Experiments	Data	entity embedding	지환 윤		
	Done	Experiments	Model	Optimizer 실험	박경택		

3. 결과 및 고찰

3.1 잘한 점

3.2 아쉬운 점

3.3 질문

3.1 잘한 점

- ✓ 대회 시작 전 세웠던 목표를 얼마나 달성했을까?

- 실험
 - [가설 수립 → 실험을 통한 가설 검증 → 사후 분석]의 프로세스를 노션에 기록하며 실험하자 !
 - 성능 향상을 위한 실험과 더불어 새로운 아이디어를 기반으로 하는 실험도 적극적으로 진행해보자 !
 - decoder로 encoder 따라잡기
 - TAPT
 - entity embedding
 - 같은 조건의 실험이라도 seed를 바꿔가며 여러 번 진행해보자 !
 - 실험 이후 적극적으로 사후 분석을 해보자 !
- github
 - git branching 전략을 활용해보자 !
 - commit message convention 합의 후 그에 맞춰 작성하자 !
- 협업
 - 프로젝트를 시작하기 전에 전체적인 timeline을 구상하여 시간을 효율적으로 활용해보자 !
 - PM 제도를 통하여 회의와 프로젝트를 원활하게 진행하자 !
 - kanban board를 활용하여 팀원들이 현재 어떤 task를 진행중인지 빠르게 확인하자 !

3.2 아쉬운 점

- 모든 실험이 아닌 일부 실험에 대한 사후 분석만이 이루어졌다.
- Decoder 계열의 모델로 encoder 계열의 모델의 성능을 따라잡아보려는 시도를 계획했으나, 실험까지 이어지지는 못했다.
- TAPT를 시도해보았는데, 논문에서 제시된 parameter 사용하지 않았고 사후 분석이 불가능했기 때문에 성능 향상으로 이어지지는 못했다.
- Entity embedding을 추가하는 실험을 진행해보았으나, 성능 향상까지 이어지지는 못했다.

End of Document

Thank You.



? 질문

- focal loss는 왜 효과적인 방법이 되지 못했는지에 대해 '쉬운 문제'와 '어려운 문제'의 경계가 모호하기 때문이라는 가설을 세워보았는데, 이것이 합리적인 가설이 될 수 있을지 궁금합니다.
- Micro f1과 AUPRC 간의 격차는 어떻게 해석해볼 수 있을지 궁금합니다.