

# Korean MRC

## With T5 and BigBird

TEAM : 2팀 나는 자연인이다.

김태웅, 설재민, 이재형, 이세린

# CONTENTS

**1**

**Project  
concept**

Introduction

-

Analysis  
process

**2**

Project  
Team

**3**

Project  
Process

**4**

**Project  
Result**

EDA

-

Model  
Selection

-

Model  
Improvement

**5**

**자체 평가  
및 보완**

**1**

**Project  
Concept**

# Project concept

## Introduction

한국어 지문을 보고 질문에 맞는 답을 생성하는  
**Machine Reading Comprehension** Project

**사용 모델** : 다양한 모델을 적용 후, 가장 성능이 좋은 모델을 선택

- Bert, RoBERTa, ALBERT ,KoELECTRA
- T5
- BigBird

T5와 BigBird를 앙상블 하였을 때 가장 좋은 성능을 보임

Framework

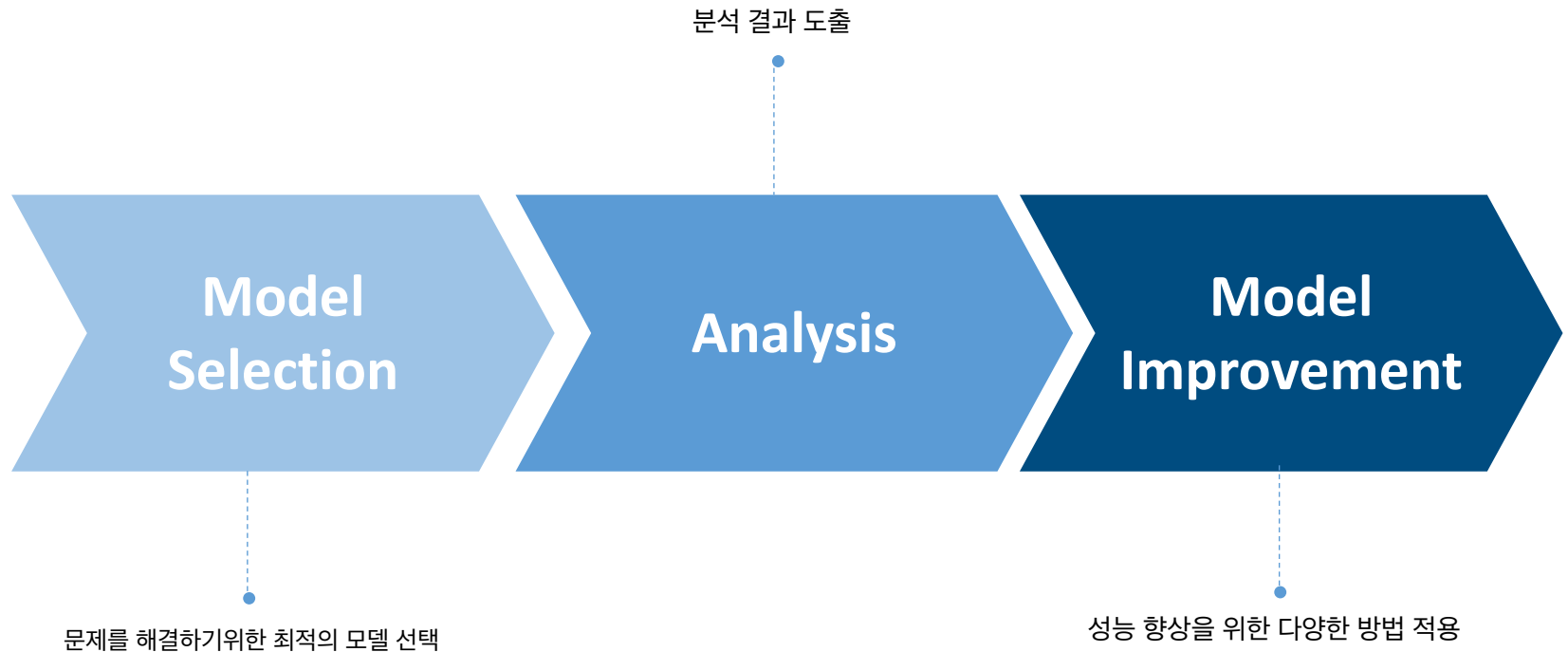


Library



# Project concept

analysis process



**2**

**Project  
Team**

# Project Team

김태웅(팀장)	Project process 설계,base line code 작성 ,T5 bigbird 모델 적용 및 앙상블, 시각화
설재민(팀원)	MRC관련 자료조사
이재형(팀원)	baseline code 수정, KoELECTRA, RoBERTa 모델 적용 시도
이세린(팀원)	Bert 및 KoBert 모델 적용 및 앙상블 적용

**3**

**Project  
process**



# Project Process

구분	기간	활동
데이터 분석	11/02 ~ 11/04	기존 Base Line 코드 및 데이터 파악
새 베이스라인 작성 & 분석	11/05 ~ 11/08	새로운 Base Line 코드 작성 및 코드 분석
모델 선정 & 적용	11/09 ~ 11/12	다양한 모델 적용 및 최종 모델 선정
모델 앙상블 적용	11/12 ~ 11/15	최종 선정된 모델을 기반으로 앙상블 적용
발표 자료 시각화 및 작성	11/15 ~ 11/16	최종 발표 자료 시각화 및 작성
총 프로젝트 진행 기간	11/02 ~ 11/16	

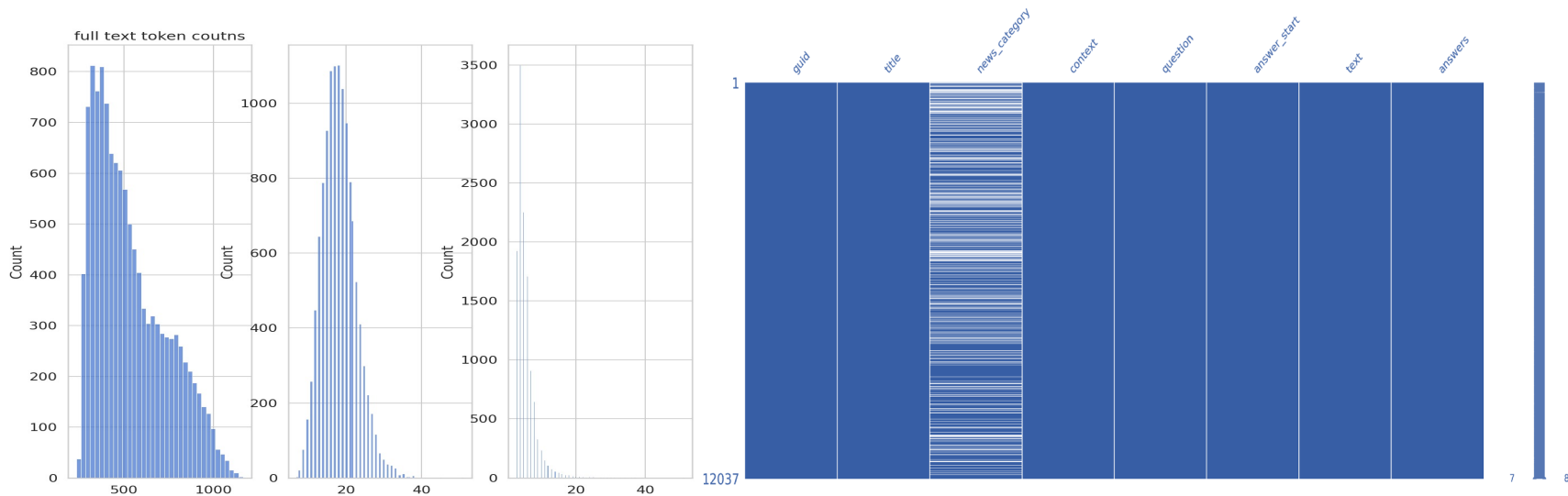
4

**Project  
Result**

# Project Result

EDA

## Train Data

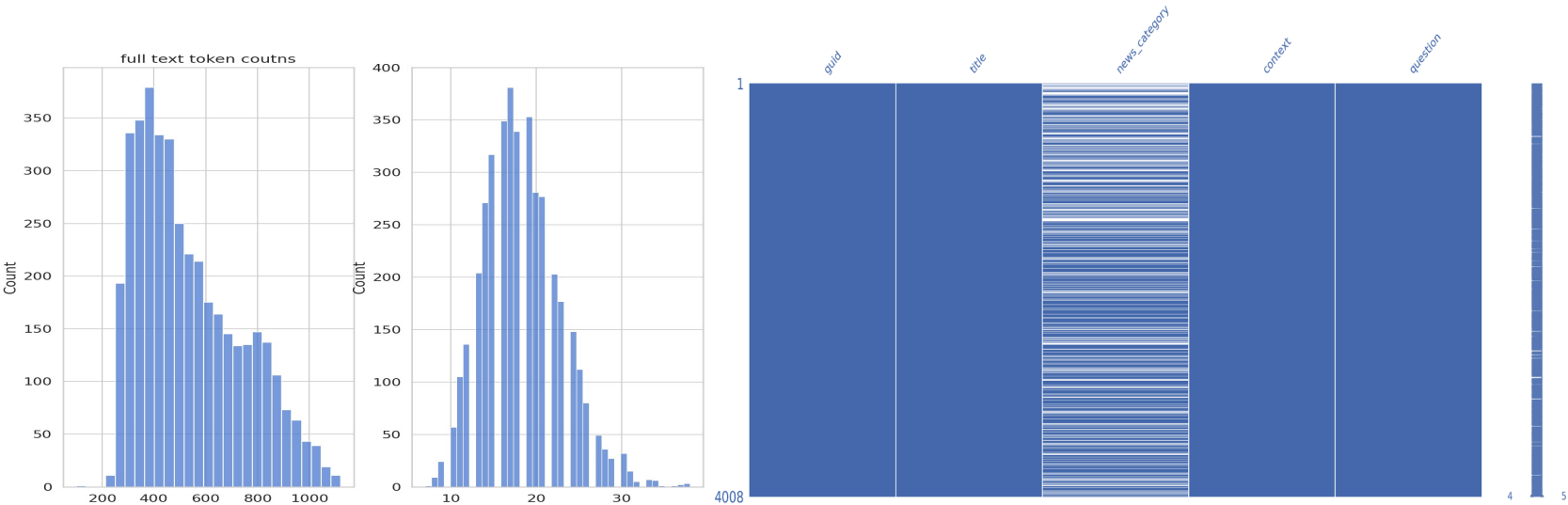


	guid	title	news category	context	question	text
count	12037	12037	8494	12037	12037	12037
unique	12037	9784	33	10434	12037	9592

# Project Result

EDA

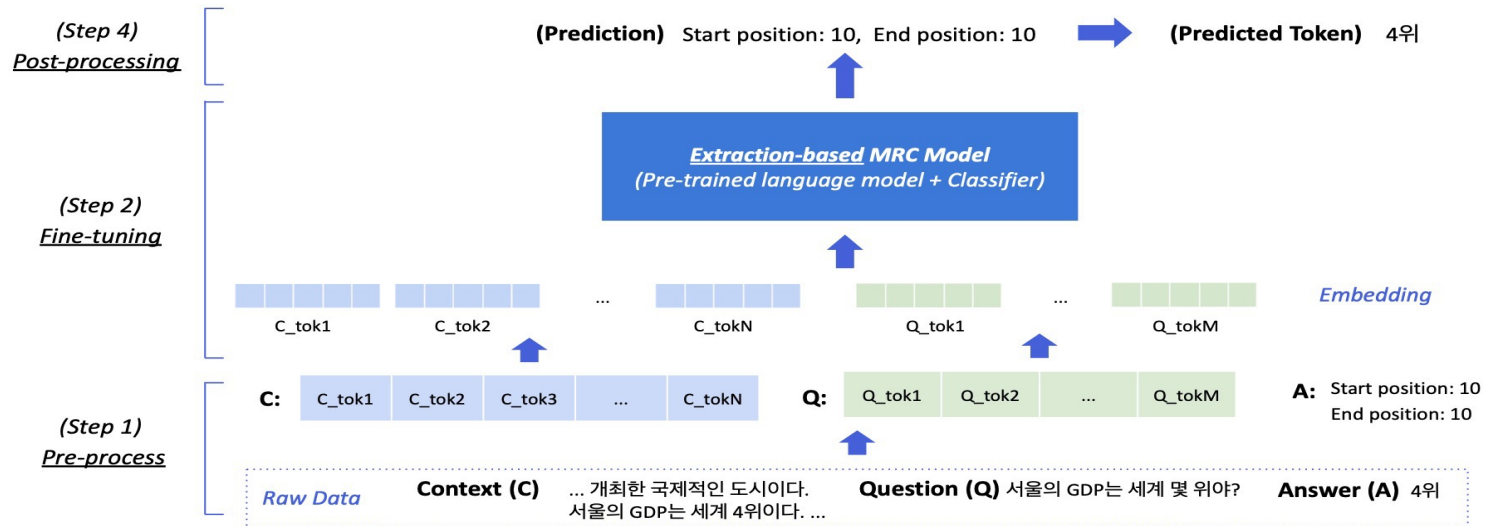
## Test Data



	guid	title	news_category	context	question
count	4008	4008	2818	4008	4008
unique	4008	3709	32	3841	4008

# Project Result

## Model selection and analysis



**문제 유형:** 지문(context) 내 답의 위치를 예측 ⇒ "분류 문제"

- 각 토큰이 정답인지 아닌지

## Loss 계산을 위한 답(Prediction)의 형태

- 지문 내 답의 위치

# Project Result

Model selection and analysis

## Extraction-based

Extraction-based MRC Model  
(Pre-trained language model + Classifier)

**(Prediction)** Start position: 10, End position: 10

**A:** Start position: 10  
End position: 10

## Post-processing

### 1. 불가능한 답 제거

- End position이 Start Position보다 앞에 있는 경우
- 예측한 위치가 context를 벗어난
- 미리 설정한 max\_answer\_length 보다 길이가 더 긴 경우

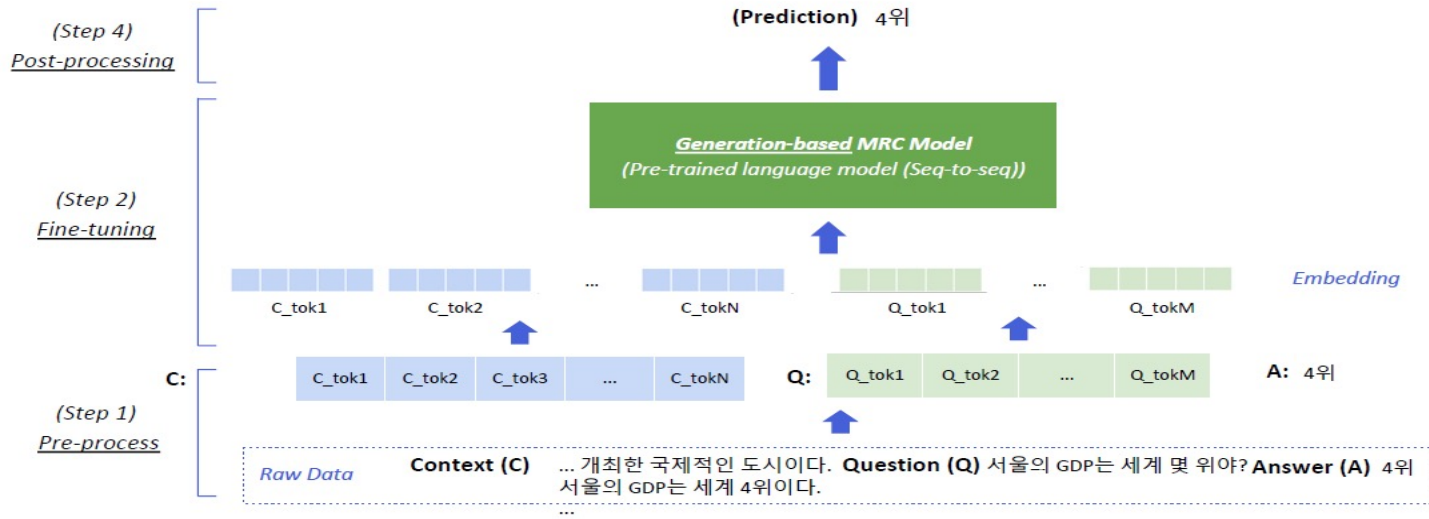
### 2. 최적의 답안 찾기

- Start/end position prediction에서 score(logits)가 가장 높은 N개를 각각 찾음
- 불가능한 start/end 조합을 제거함
- 가능한 조합들을 score의 합이 큰 순서대로 정렬함
- Score가 가장 큰 조합을 최종 예측으로 선정함

시작과 끝에 해당하는 contextualized vector(embedding)를 scalar value로 내보냄

# Project Result

Model selection and analysis



**문제 유형:** 주어진 지문과 질의를 보고 답변 생성 ⇒ "생성 문제"

- 지문 내 정답이 있든 말든 상관 없음

**Loss 계산을 위한 답(Prediction)의 형태**

- Free-form text 형태

# Project Result

Model selection and analysis

## Generation-based

Generation-based MRC Model  
(Pre-trained language model (Seq-to-seq))

(Prediction) 4위

4위

## Post-processing

**Greedy Search:** decision making을 early stage에서 하므로 초반에 틀리면 다 틀릴 지도 모른다는 단점

**Exhaustive Search:** 모든 가능성을 다 봄 → 가능한 가짓수가 time step에 비례하므로 문장이 조금만 길어져도, vocab size가 조금만 커져도 불가능

**Beam Search:** exhaustive search를 하되, 각 time step마다 가장 점수가 높은 top-k만 유지 → 가장 많이 쓰는 방식

- Extraction-based MRC는 정답의 위치를 정확히 특정해야 했는데 Generation-based MRC는 그럴 필요 X  
→ 정답 그대로 넘겨주면 됨
- 전체 sequence의 각 위치마다 모델이 아는 모든 단어들 중 하나의 단어를 맞추는 분류 문제



# Project Result

Model selection(Bert) and analysis

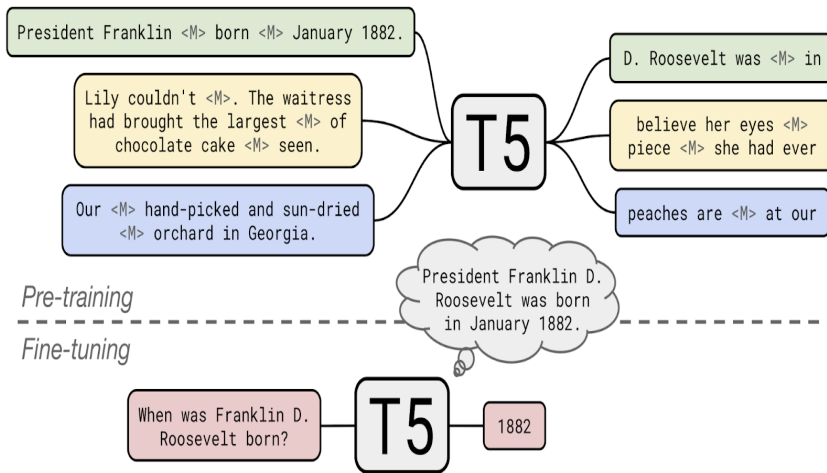
Model	levenshtein
Bert	37.393
KoBert	16.271
KoBert(output_MAX_LEN 10)	9.022

## Test Review

- Bert보다, Bert 모델을 한국어 기반으로 제작한 KoBert가 더 좋은 성능을 보임
- Train 데이터의 answer 평균 길이가 7~8정도임을 감안하여 예측 값의 최대값을 제한하자 성능이 향상됨
- Pretrain model(bert)의 input max\_len이 512로 제한되어 있어 주어진 긴 시퀀스의 답을 예측하는데 한계가 있다.

# Project Result

Model selection and analysis



T5



## Big Bird: Transformers for Longer Sequences

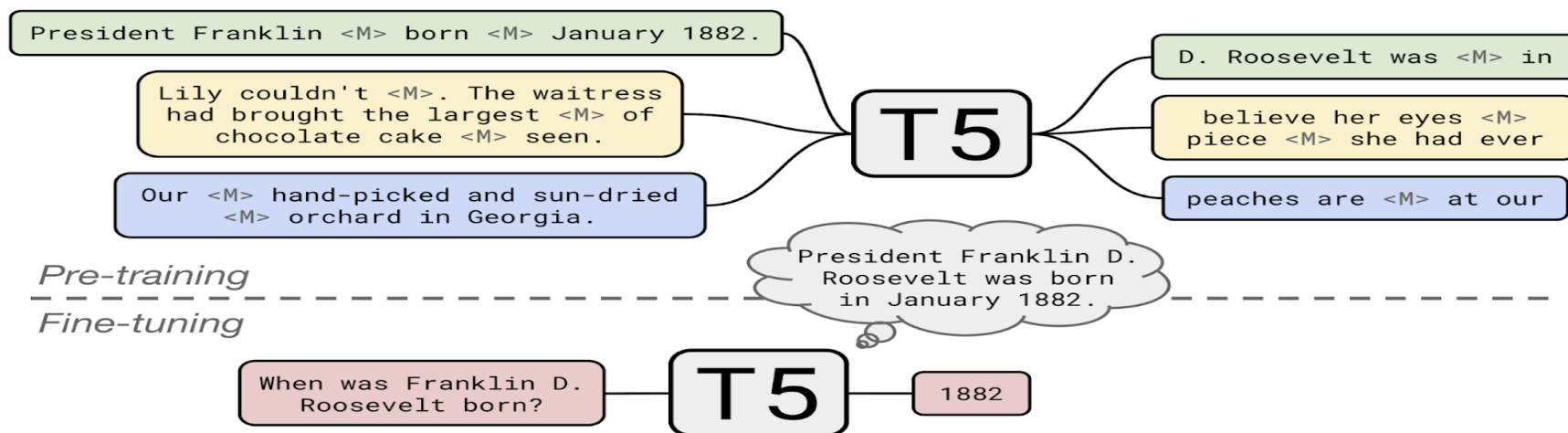
Manzil Zaheer, Guru Guruganesh et al.

Google

Bigbird

# Project Result

Model selection(T5) and analysis



구글에서 제안한 T5 구조는 상당히 특이한 형태로 문제들을 기술합니다. 기존의 모델들이 배치 단위에서 실 훈련 데이터를 변경하며 다양한 문제를 모델에 투입하였다면, T5는 모든 문제를 문장 형태로 추상화 한 다음에, 그 추상화된 문장을 푸는 것을 훈련시킨 구조입니다.

# Project Result

Model selection(T5) and analysis

- **version\_1**

- data : groom competition에서 주어진 데이터
- model : T5
- parameter : max\_length -> 512, learning\_rate -> 0.0001, epoch -> 5, batch\_size -> 4

- **version\_2**

- data : groom competition에서 주어진 데이터 + AI hub(25만개)
- model : T5
- parameter : max\_length -> 512, learning\_rate -> 0.0001, epoch -> 5, batch\_size -> 4

## Test Review

- AI Hub의 MRC 데이터 25만개를 추가하여 학습했을 때 성능이 떨어짐.
- Colab의 메모리 부족으로 batch\_size의 크기는 4가 최대이며 batch\_size가 증가할 수 록 점수가 개선됨
- Learning는 하이퍼파라미터 튜닝을 진행하지 못함
- **Version\_1을 기준으로 Post processing**

# Project Result

Model selection(T5) and analysis

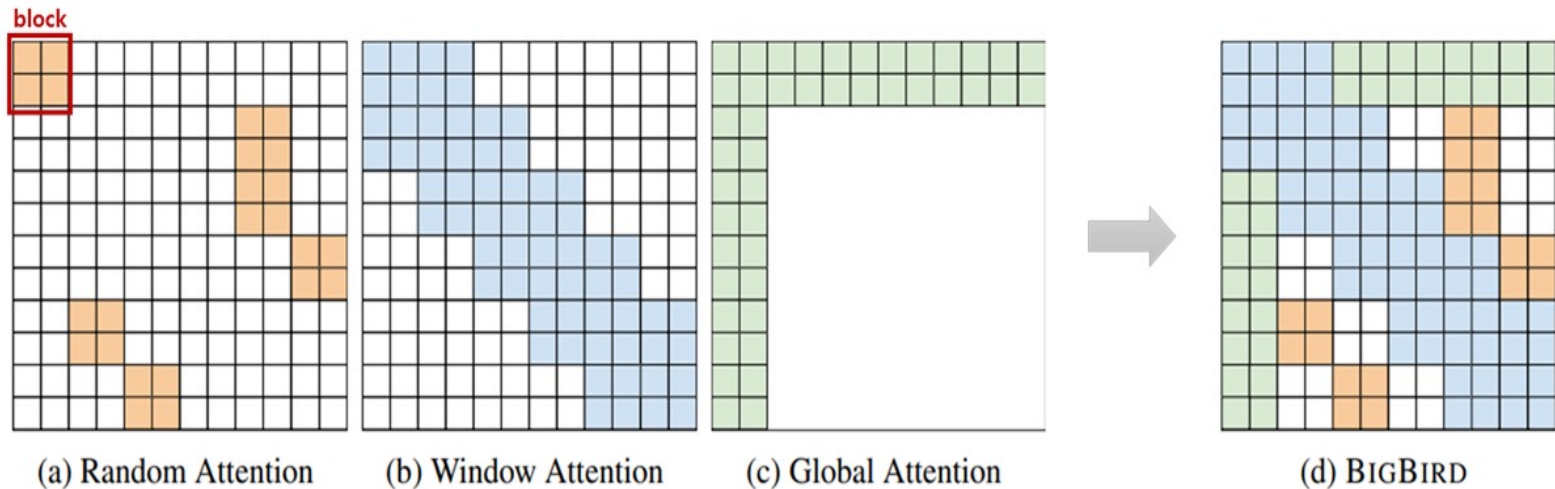
Model	num_beam	max_lenght	levenshtein
T5-base	1	32	3.591
T5-base	5	32	3.558
T5-base	5	20	3.490
T5-base	10	20	3.548

## Test Review

- Num\_beam은 5일 때 가장 성능이 좋음.
- Output의 max\_lenght는 20일 때 성능이 좋음.
- Num\_beam :5 , max\_length : 20

# Project Result

Model selection(BigBird) and analysis



BigBird는 Transformer architecture를 기반으로 합니다. 즉 multi-head self-attention과 feed-forward network로 구성된 layer를 여러겹 쌓아서 만든 구조이며, 다만 self-attention layer에서 full-attention이 아닌 sparse attention으로 연산하는 것이 차이점입니다.

# Project Result

Model selection(T5) and analysis

- **version\_1**
  - data : competition에서 주어진 데이터
  - model : BlgBird
- **version\_2**
  - data : competition에서 주어진 데이터 + AI hub(3만개)
  - model : BlgBird
- **version\_3**
  - data : competition에서 주어진 데이터 + AI hub(5만개)
  - model : BlgBird

## Test Review

- Bigbird 모델의 3가지 버전의 하이퍼 파라미터는 동일함
- parameter : max\_length -> 1024, learning\_rate -> 0.0005, epoch -> 5, batch\_size -> 4
- **Version\_1, Version\_2에서 성능이 우수함**

# Project Result

Model selection(BigBird) and analysis

Model	public socre	private socre
BigBird (version_1)	3.79725	3.44698
BigBird (version_2)	2.77916	2.59043

## Test Review

- AI\_hub의 데이터 3만개를 추가한 버전에서 성능이 가장 높게 나왔다.
- Version\_3(데이터 5만개 추가) validation data에서 성능이 매우 낮게 도출됨
- Bigbird 모델이 주어진 데이터의 긴 문장을 잘 파악하는 것으로 보임



# Project Result

Model selection(Esemble) and analysis

Submission and Description	Private Score	Public Score
<a href="#">final_MRC_v2.csv</a> 8 minutes ago by <a href="#">Tae Woong</a> <a href="#">add submission details</a>	2.55634	2.76855

## Test Review

- T5(version\_1), bigbird(version\_1), bigbird(version\_2)를 Esemble한 결과 성능이 개선됨
- 결과값은 다수결로 채택
- 3 모델의 결과값이 다른 경우 bigbird(version\_2)의 모델 결과값을 선택

5

자체 평가  
및 보완

# 자체 평가 및 보완

## 다양한 모델 활용

다양한 모델을 문제에 적용해 보면서 각 모델의 특성에 대해서 파악하고 문제에 적합한 최적의 모델을 찾는 경험을 했다.

## Gpu의 메모리 용량 제한

Gpu의 메모리 용량의 한계로 input data의 max\_length의 길이를 늘리는데 한계가 있고, Batch\_size의 크기에도 제한 사항이 있다.

## HyperParameter Tuning

Learning\_rate, epoch, Regularization 등의 파라미터 튜닝을 다양하게 진행하지 못했다.

## Esemble 적용

2가지 모델 앙상블 결과 성능이 개선되었고, 더 다양한 모델을 Ensemble을 적용 경우 성능 향상을 기대할 수 있다.

**THANK  
YOU!**