

Chatbot 알고리즘

BERT & Bagging

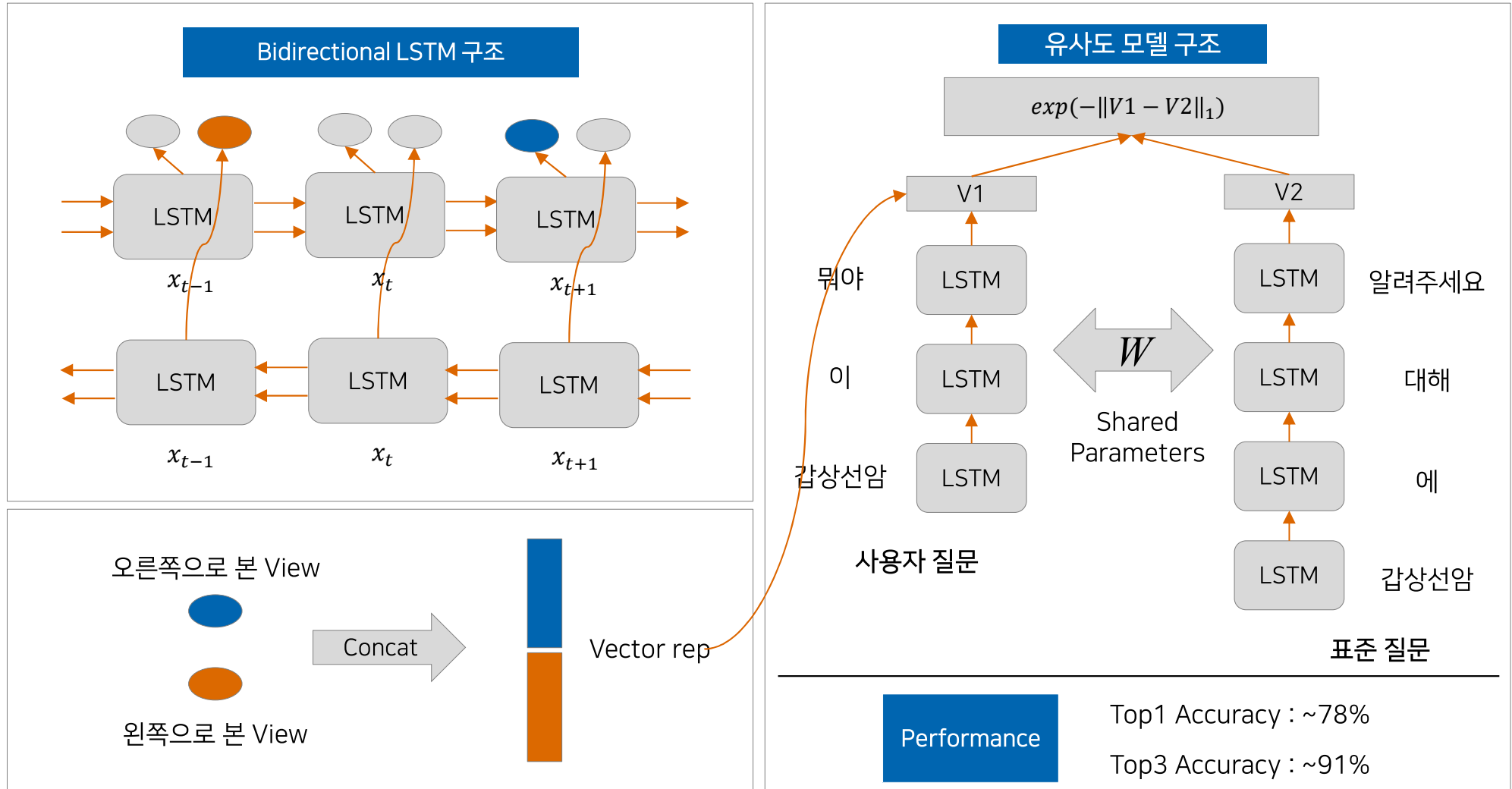
2020. 09. 24.

Section 1

LSTM Siamese Similarity & Classification Model

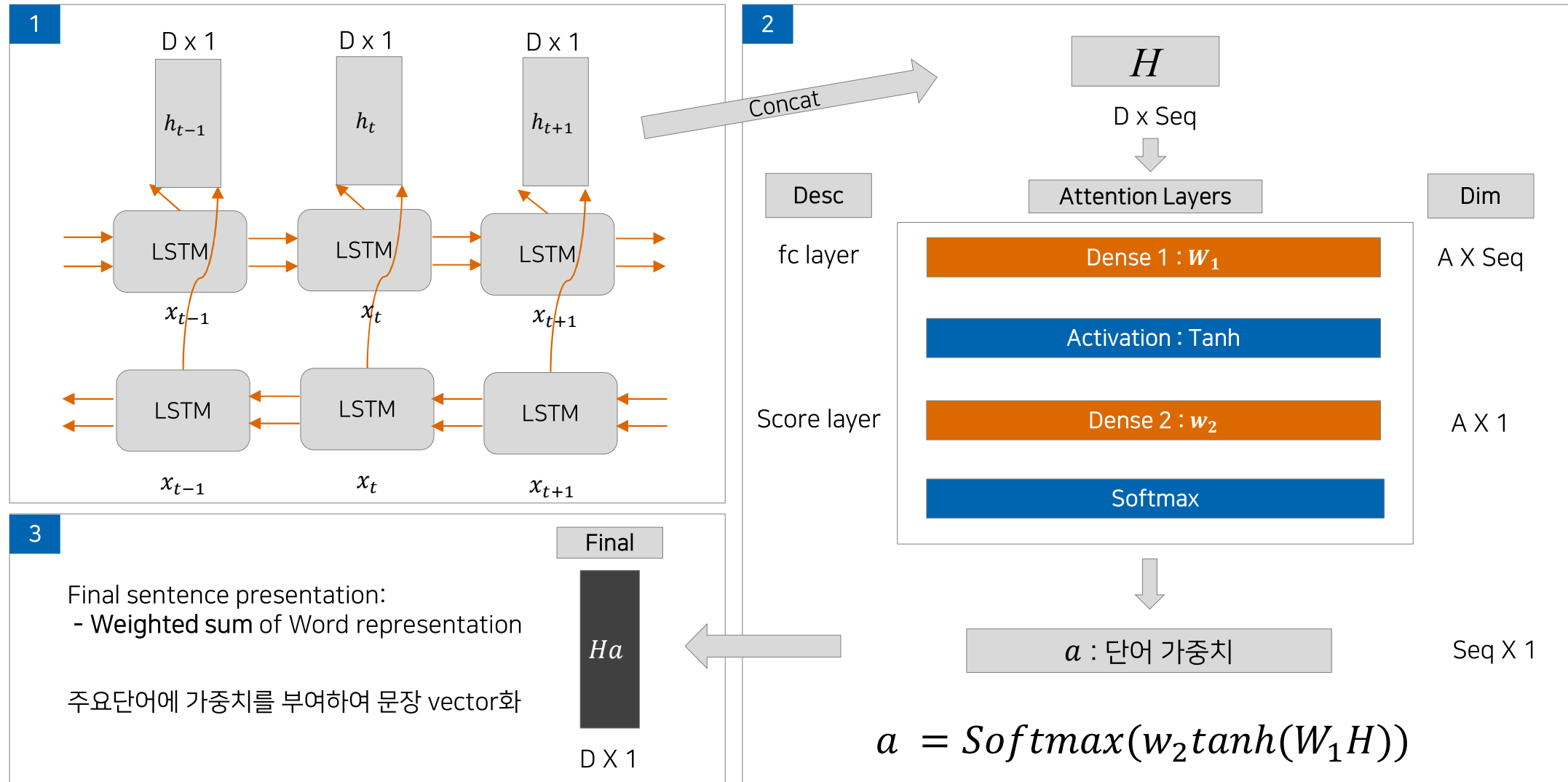
Siamese Network

예전 모델에서 사용하는 알고리즘은 BiLSTM 모델이며 문장을 양쪽에서 processing 함
각 방향에서의 최종 hidden-layer들을 concat하여 문장의 vector representation으로 구성



Attention Mechanism

모델을 훈련하면서 Task를 수행할 때 중요하게 사용된 단어에 attention을 더 주어 높은 가중치 부여
사용 모델에서는 오른쪽 왼쪽 view에 각각 다른 attention을 주고 sum을 하여 vector화함 (순서)



제한점 개선 사항 - Attention의 문제 : Masking

몇몇 class의 경우 일정 단어가 없으면 맞추질 못함 (단어의 의존도가 너무 강함)

Masking으로 훈련 시 무작위로 단어를 삭제 후 input을 하면 단어의 의존도가 줄어드는 것을 알 수 있음

Input

Ex) 하악골 절제술

Answer

구강암 하악골 절제술은 어떻게 진행되나요

Non-masked model

유사도

1. 광범위 외음부 절제술에 대해 알려주세요

27%



Class를 맞추기 위하여 한 단어에 너무 많이 집중함 (구강암)

Masked model

유사도

1. 구강암 하악골 절제술은 어떻게 진행되나요

72%



- 한 단어에 너무 의존하지 않음
- 하지만 여러 단어를 쓰려하다 보니 노이즈에 조금 더 취약
- Testing Accuracy : 94%

Masking Process

('구강암', 'Noun'), ('하악골', 'Noun'), ('절제', 'Noun'), ('술', 'Noun')



Delete

('구강암', 'Noun'), ('하악골', 'Noun'), [delete], ('술', 'Noun')



Feed

Model

- Training set size 증가
- 단어 의존도 reduction

개요 - 기존 모델 Summary

3가지 모델이 있으며, 각 모델의 장단점이 있을 것으로 예상

Attention + Siamese (Siam)

장
점

- 표준 답변 추가 시 따로 훈련 필요 x
- 문장 내 모든 단어를 적절히 사용함

Top1 : 92%
Top3 : 96%

Multi-Attention Classification (NoMask)

- Clear한 단어선택 (노이즈 제거)
- 높은 Test performance (96%)

Top1 : 96%
Top3 : 98.3%

Multi-Attention + Masked model (Mask)

- Mask가 없는 모델 보다 더 많은 단어를 봄
- Class에 한 단어 의존도를 완화
- 오타가 있는 경우에도 알아들음

Top1 : 94%
Top3 : 97.8%

단
점

- 비교적 낮은 Test performance (92%)
- 너무 많은 훈련데이터 필요
- 명확하지 않은 attention을 사용하다 보니 다른 모델에 비해 큰 노이즈 문제

- 가끔 엉뚱한 단어에 Attention이 들어가 방해가 되기도 함
- 짧은 문장에서는 Attention이 큰 의미가 없어짐
- 순서 의존도가 아직 남아 있음

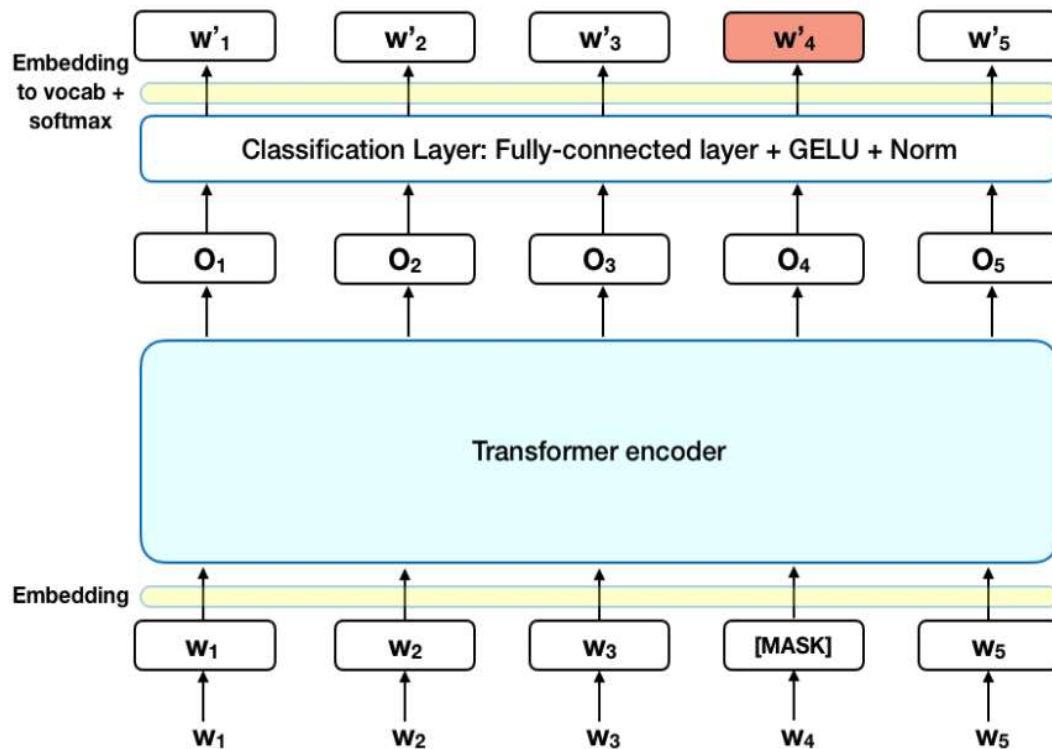
- Wrong Attention
- 노이즈 문제가 다시 생김
- 비교적 낮은 퍼포먼스 (94%)

Section 2

BERT Classification Model

모델 추가 - BERT : Bidirectional Encoder Representation for Transformer

사전 학습 기반인 SKT-KoBERT 모델을 Fine tuning 하였으며, 기존에 문제 되었던 (순서적 의존도 등) 부분에서도 좋은 performance를 보여줌



Pre-Training

Masking Mechanism

- Deeper Bi-directionality

Attention Mechanism

- Information selection(attend할 단어 선택)으로 인한 Noise reduction
- Self-dot product attention + Multi-head attention

Next-sentence Prediction

- Context recognition

Fine-Tuning

- CLS token for classification
- Question Answering task & more

Top1 Performance : 96.9%

Section 3

Model Evaluation & Setting up Bagging Strategy

Model Aggregation Strategy – Correct model selection for each instance

적시에 모델을 사용하기 위해서는 각 모델들이 정답을 맞췄을 때의 특징 파악이 필요 (Input/Output)
특징 분석 이후, 어떤 모델의 결과를 내보낼지 결정하는 전략을 설정 할 수 있을 것으로 예상

Input Feature

- 문장 길이 (Token length)
- 문장 구성 (명사, 조사 etc. 개수)

Indicator

- Noisiness
- 순서 변화
- Limited info

Problems

- 문장길이를 모델선택에 애매함 (Tukey HD)
- Tokenizer가 정확하지 않음 (OKT)
- BERT는 다른 Tokenizer 사용 (WPE)
- Input feature들로 모델을 선택하는 것은 힘들 것으로 보임

Model

BERT

Siam

Mask

NoMask

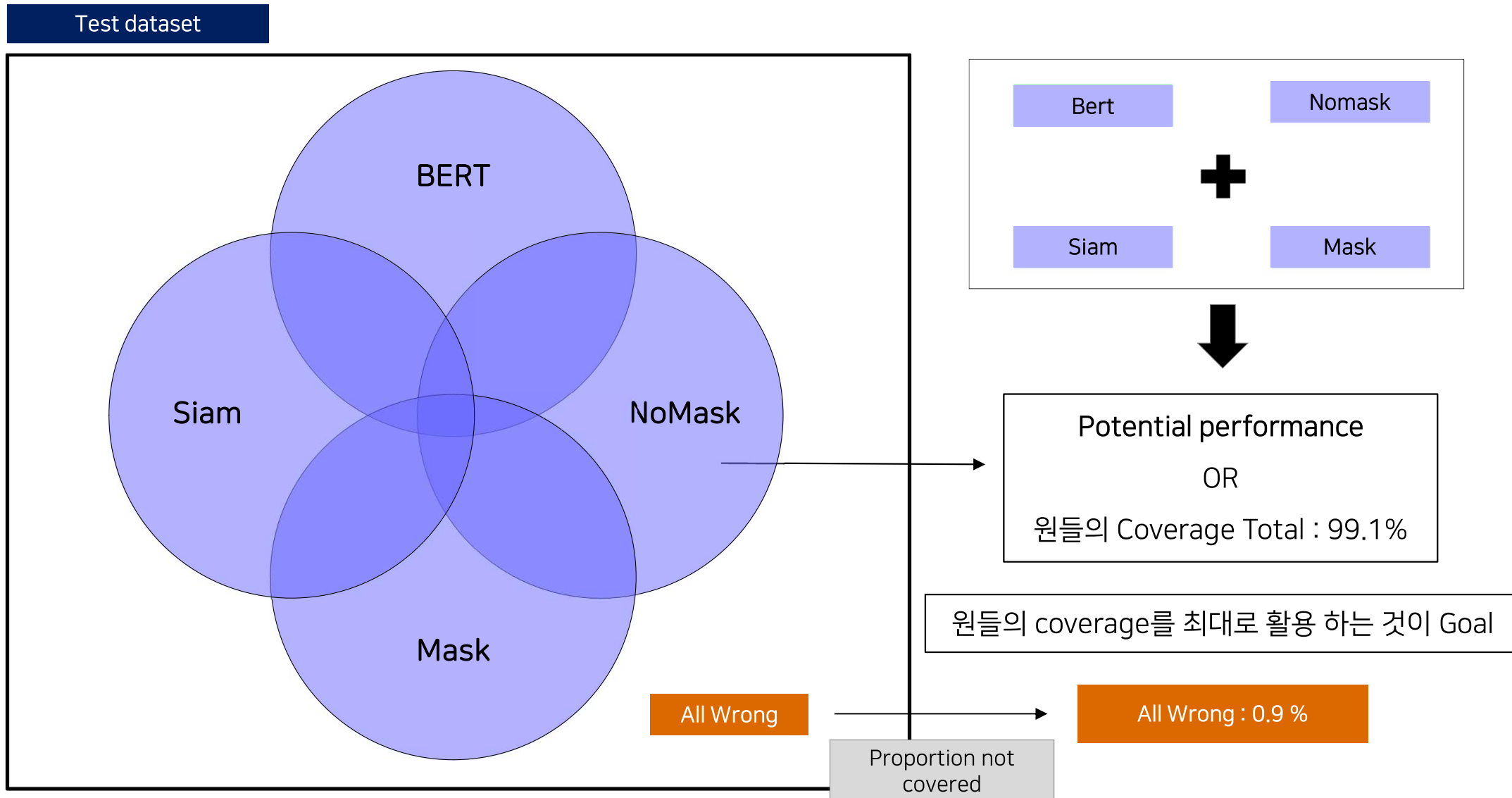
Output features

- Top1 유사도
- Top1 유사도 - Top2 유사도

Output feature인 유사도를 분석하여 사용자 최종 output form을 정하는게 적절할 것으로 예상 (가설)

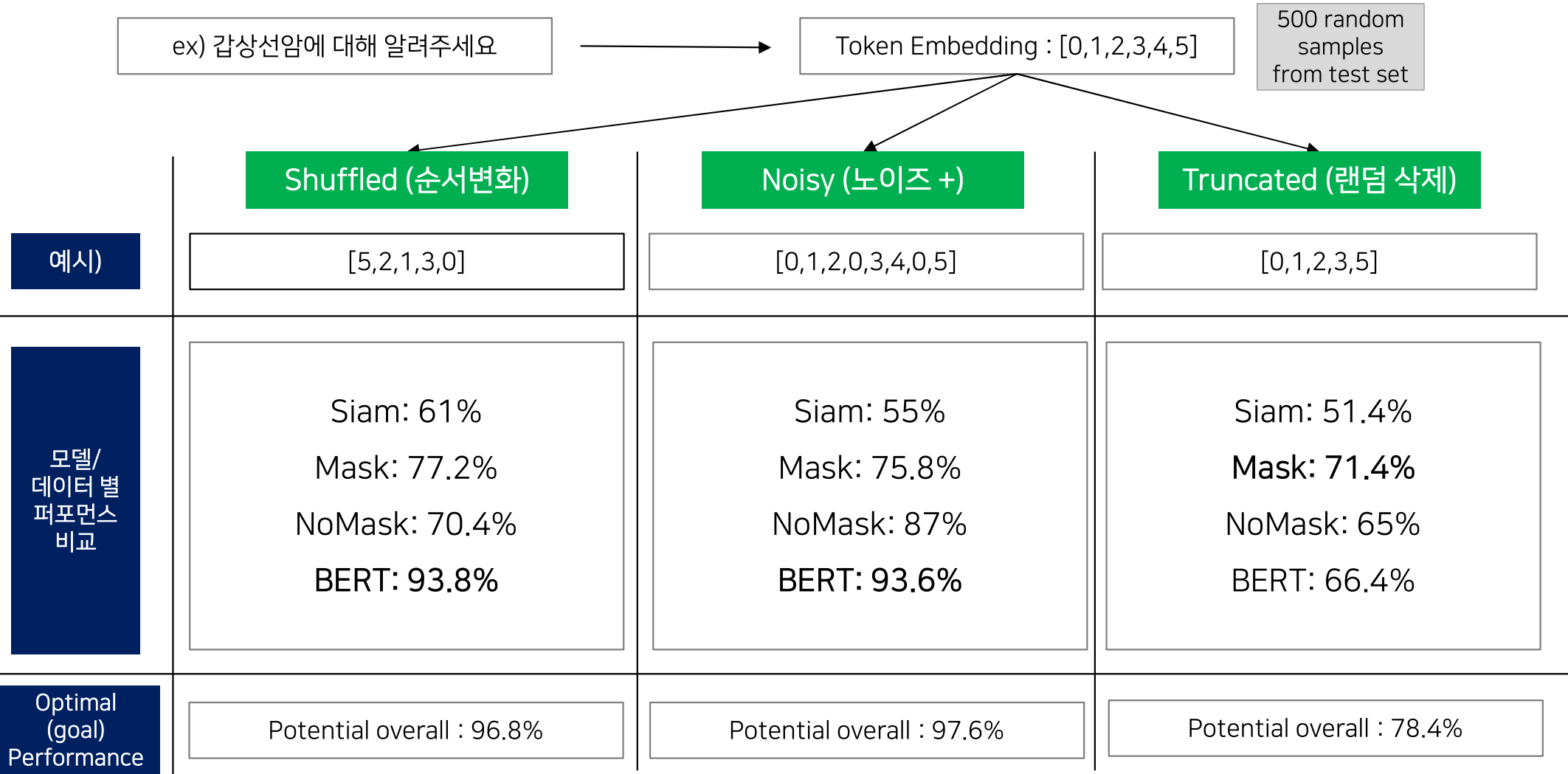
Model Aggregation Strategy - Goal

총 4개의 모델의 강점을 적시에 활용하면 performance를 향상 시킬 수 있을 것으로 예상



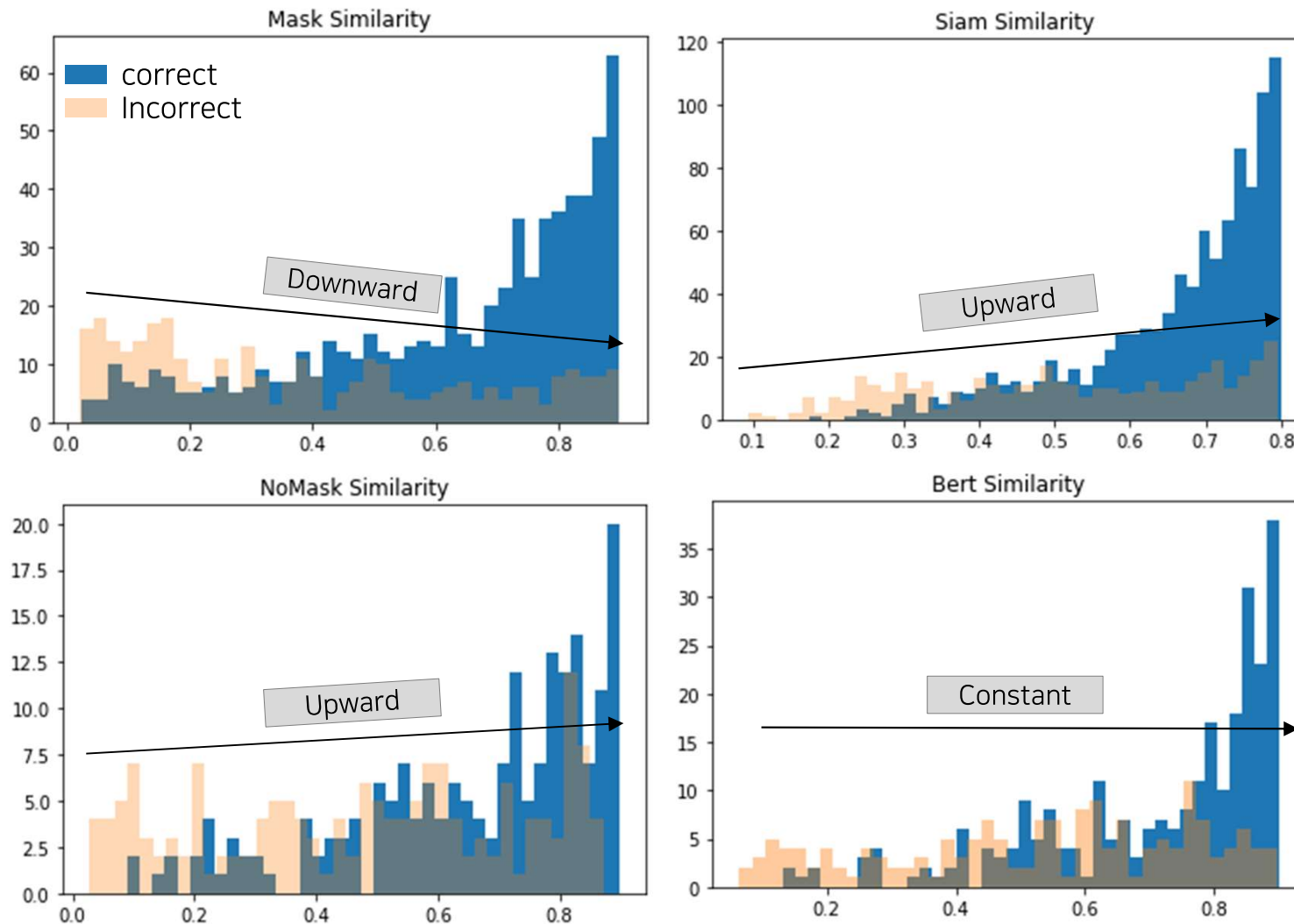
Model Aggregation Strategy – Simulated data

현재 Test set은 너무 정형화 되어 있음
새로운 데이터를 받았을 때의 performance를 예측하기 위해서 3가지 종류의 data를 구성함 (500 ea)



Model Aggregation Strategy – 유사도 분석 (Original test set)

Top1 유사도 분포 확인 결과, 모든 모델이 유사도가 높을수록 맞췄을 가능성이 높아지는 것으로 보임
맞추지 못한 Case의 경우 유사도가 증가 할수록 각각 다른 trend를 보여줌



Average Similarity

Mask

Wrong Case sim : 46%
Correct Case sim : 95%

NoMask

Wrong Case sim : 66%
Correct Case sim : 99%

Siamese

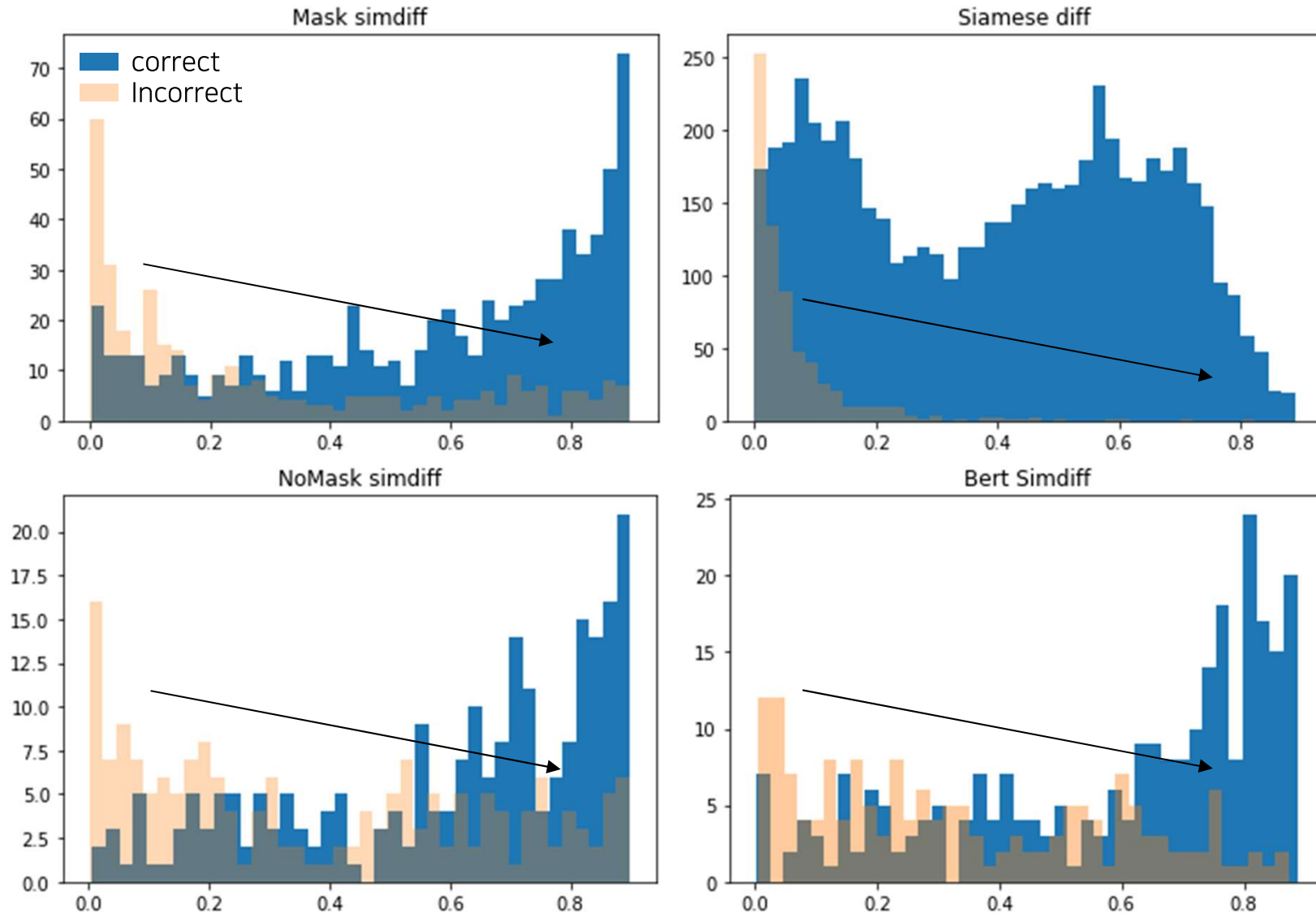
Wrong Case sim : 66%
Correct Case sim : 87%

Bert

Wrong Case sim : 59%
Correct Case sim : 97%

Model Aggregation Strategy – Top1 Top2 유사도 차이 분석 (Original test set)

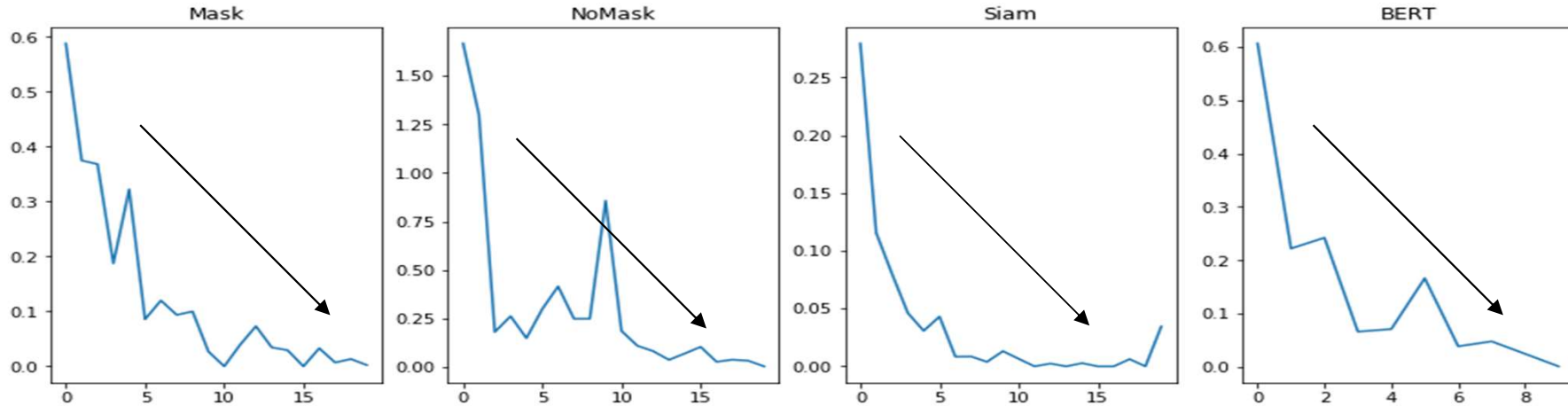
Top1 Top2 두 유사도 차이의 분포 확인 결과, Top1 유사도 분포보다 더 확연한 downward trend가 틀린 Case에서 확인됨



Model Aggregation Strategy - 유사도 분석 (Original test set)

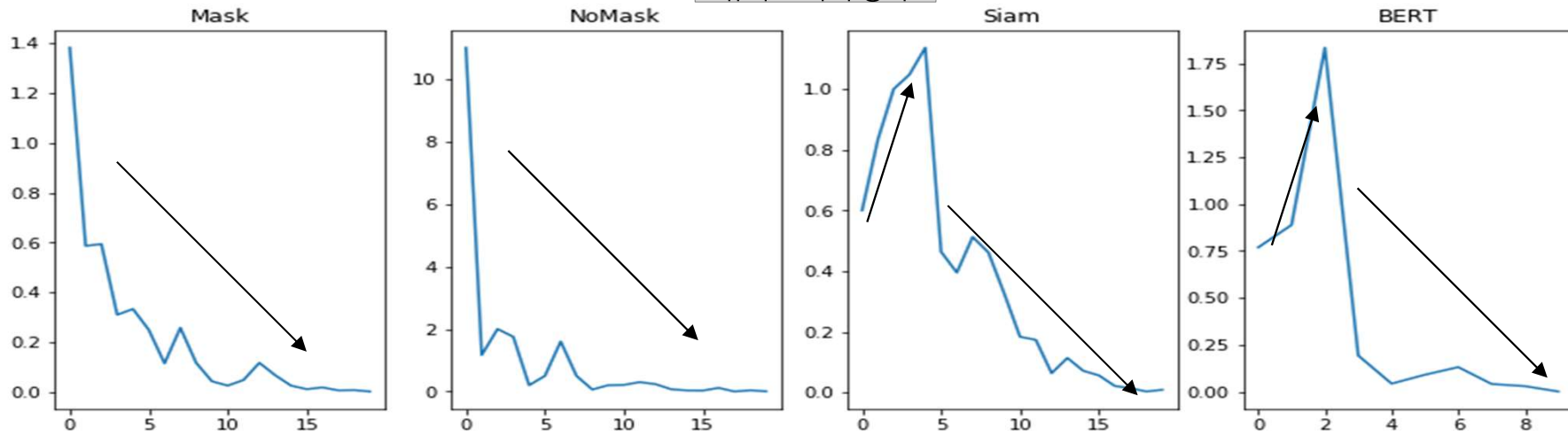
유사도 차이 증가에 따른 (틀린 case 수: 맞은 case 수) 의 ratio가 전반적으로 떨어지는 것으로 보임
유사도가 증가 할 때도 비슷한 트렌드를 보이거나 유사도 차이에서 더 뚜렷한 트렌드를 보임

유사도 차이



유사도 차이 증가

유사도

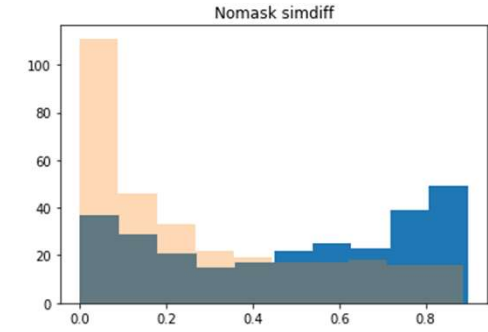
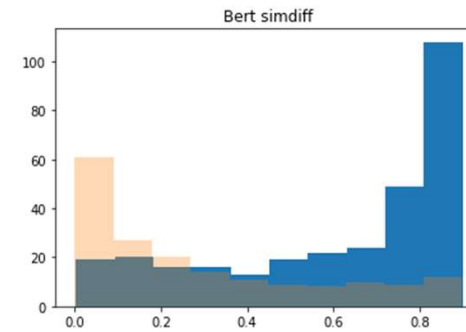
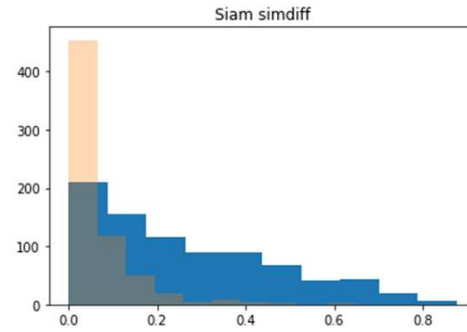
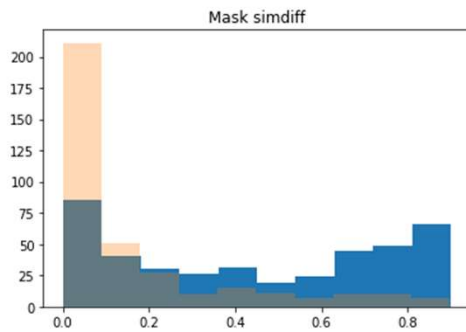


유사도 증가

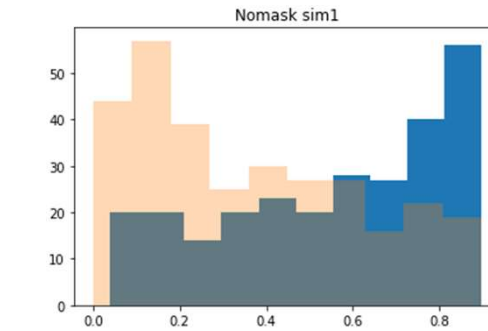
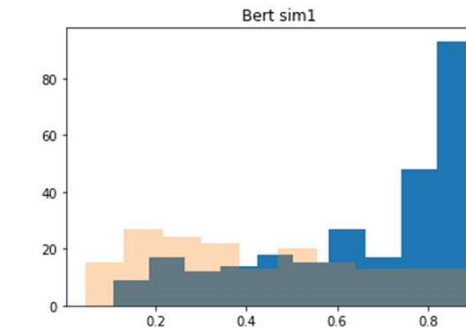
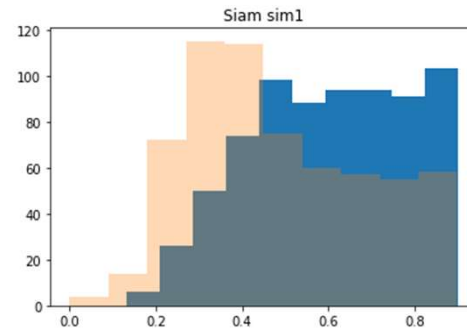
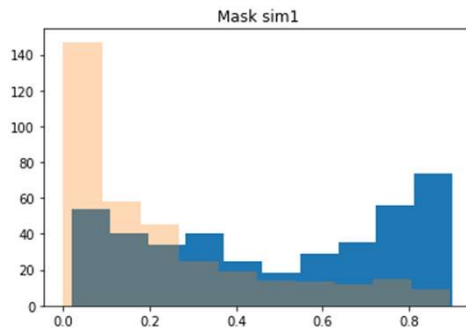
Model Aggregation Strategy – 유사도 분석 (Generated test set)

Generated data (Shuffled, Noisy, etc.) 에서도 같은 패턴을 보이는 것으로 보여짐

유사도 차이

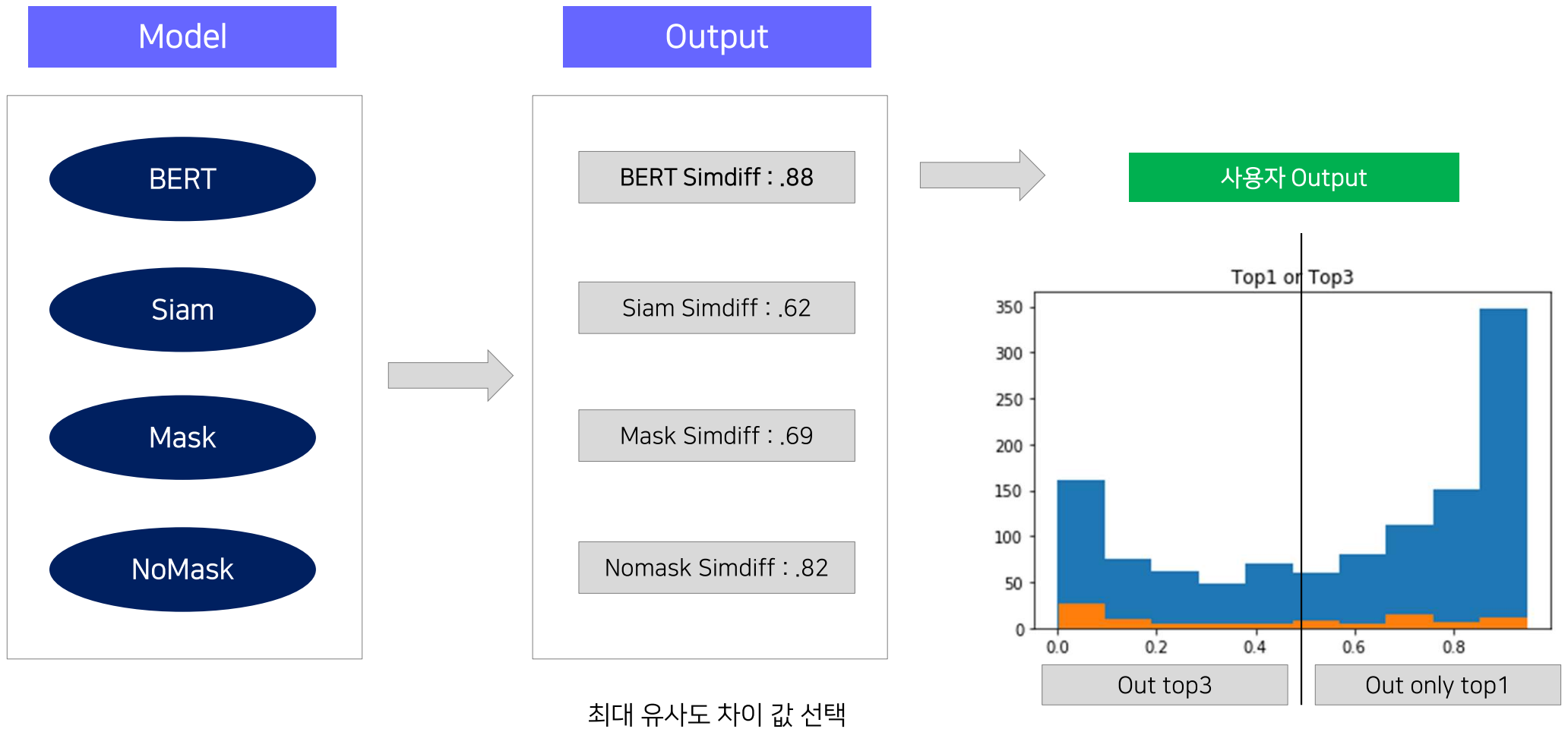


유사도



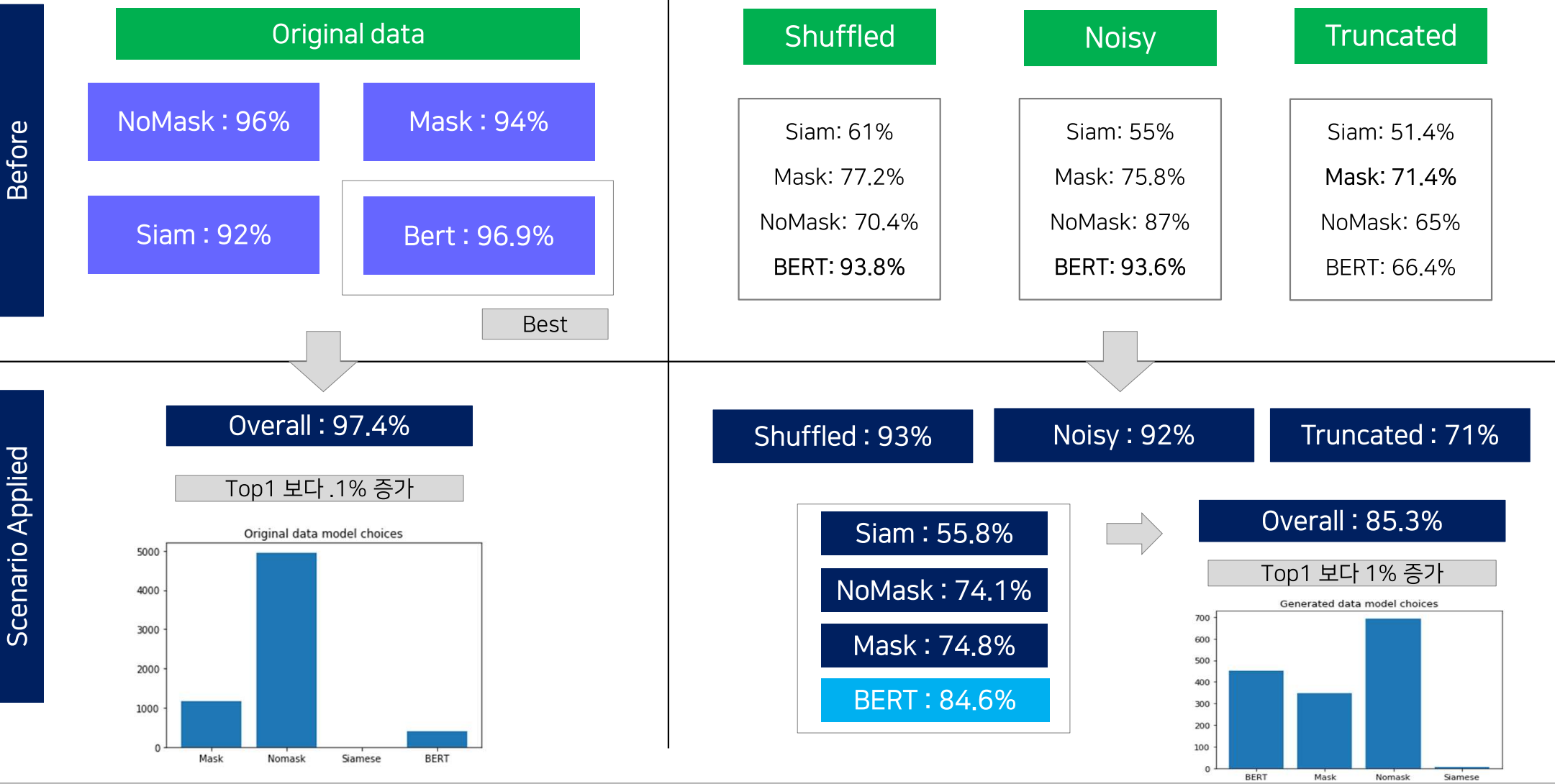
Model Aggregation Strategy – Model selection

유사도의 차이가 클수록 높은 확률로 정답을 맞출 수 있고, 유사도 차이를 정답에 대한 confidence로 정의
가장 높은 confidence를 갖는 outcome을 model output으로 설정



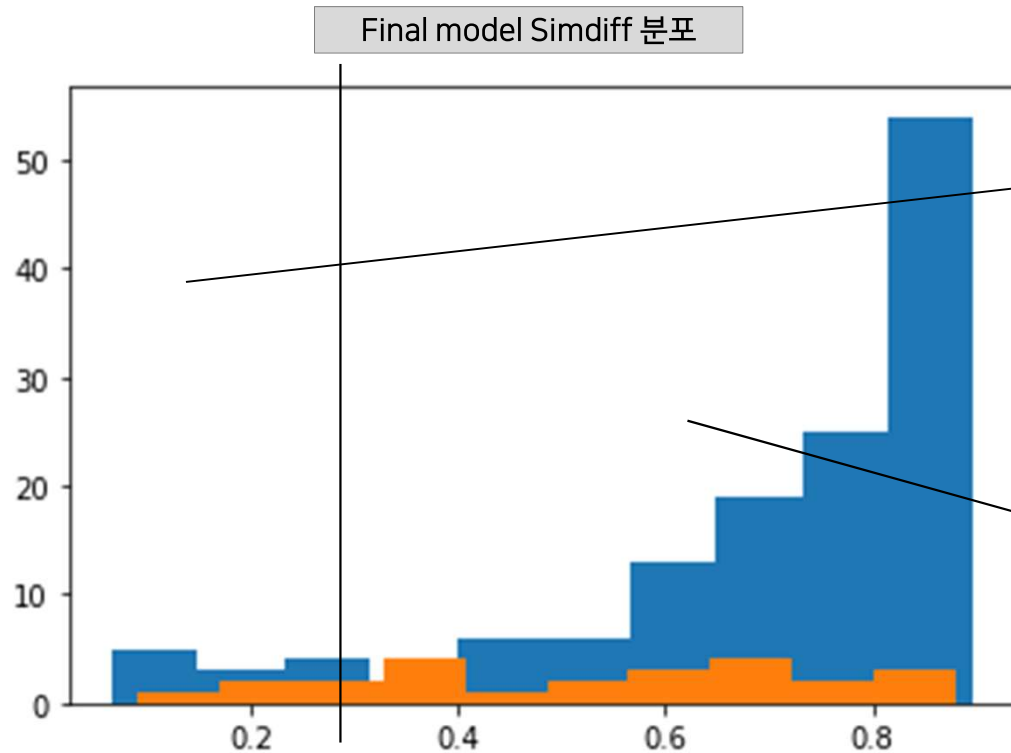
Model Aggregation Strategy - Performance

전반적인 Improvement가 있었으나 Truncated case에서 Drop이 있음
 모든 모델을 이용했을때 가장 좋은 퍼포먼스가 나오는 것으로 확인



Other Addition - 상담원 연결

특정 threshold 이하의 경우 틀린 비율이 맞춘 비율과 비등한 것을 볼 수 있음
Threshold 이하의 경우 상담원과 연결 시키거나 이해를 못했다는 문구를 내보낼 수 있음



Answers not even in top 3

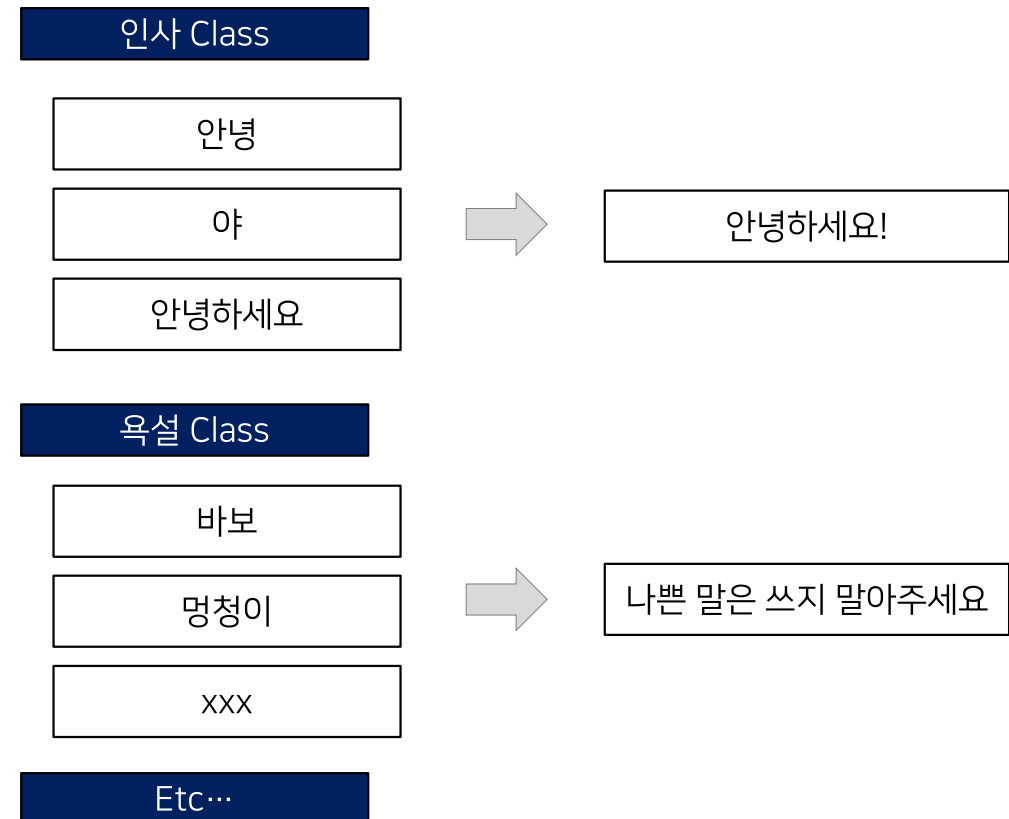
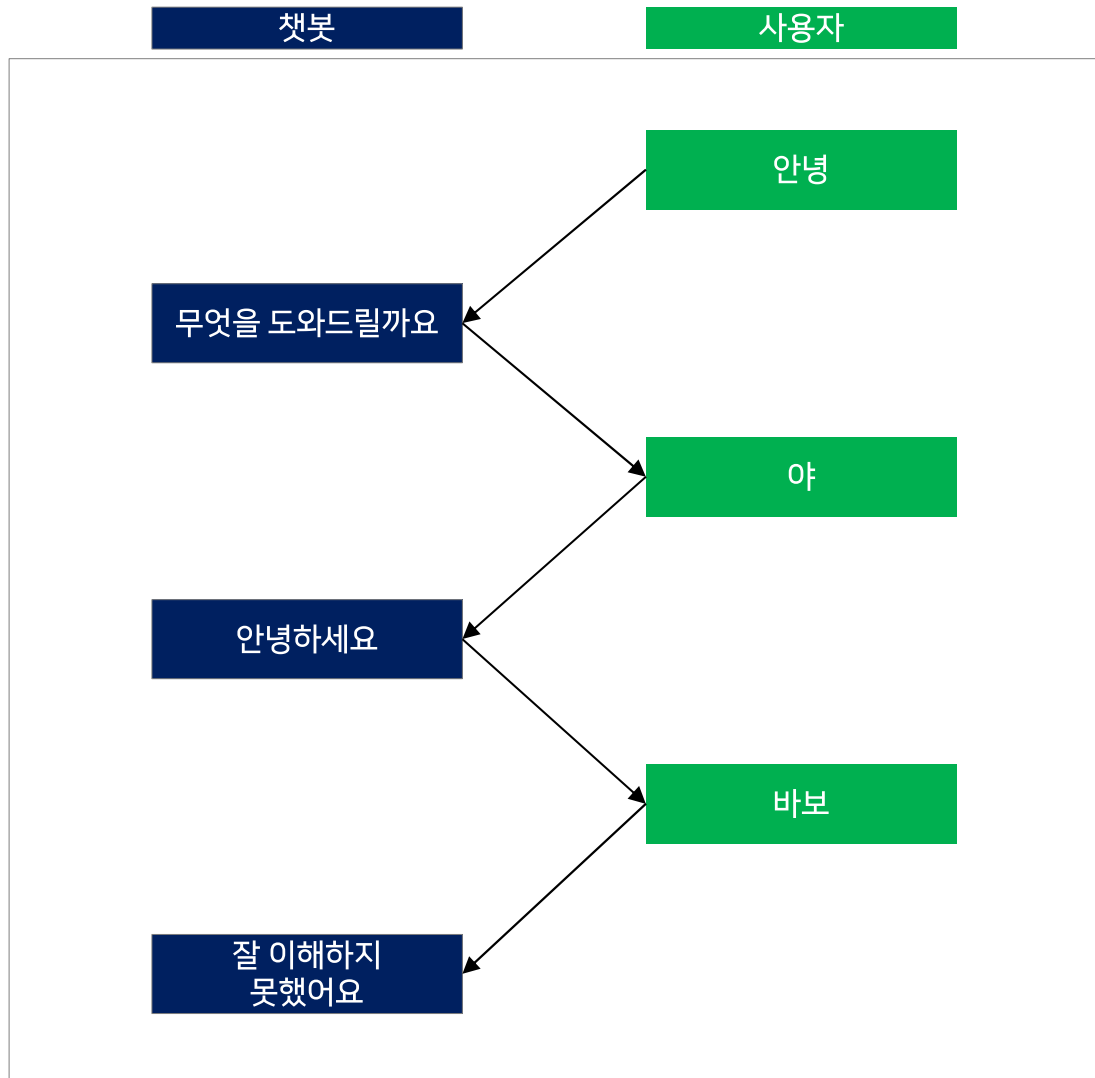
Answers in top3

**제가 잘 이해하지 못했어
요. 회원서비스 관련 질문
을 쉽게 설명해주시면 더
좋은 대답을 할 수 있어요.

Original 답변 제공

To-be - Irrelevant cases

기본적인 인사나 대화 등을 추가 할 수 있음



End of Documents