

# 계층적 loss 및 representation과 레이블 임베딩을 이용한 논문 문장 수사학적 분류 모델

국내 논문 문장 의미 태깅 모델 개발

INU\_DILAB

강병하, 국희진, 김영화, 윤세휘

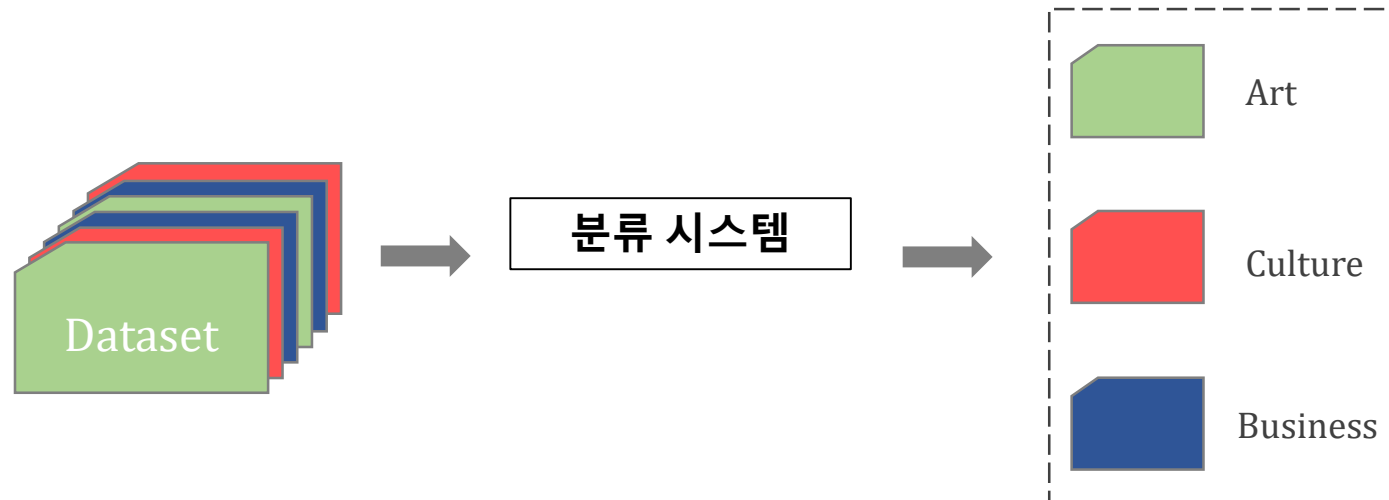
# 목차

- 1** 프로젝트 개요
- 2** 활용 데이터
- 3** 모델 개발 방법
- 4** 실험 및 평가
- 5** 활용 계획 및 기대효과

# 문제 정의

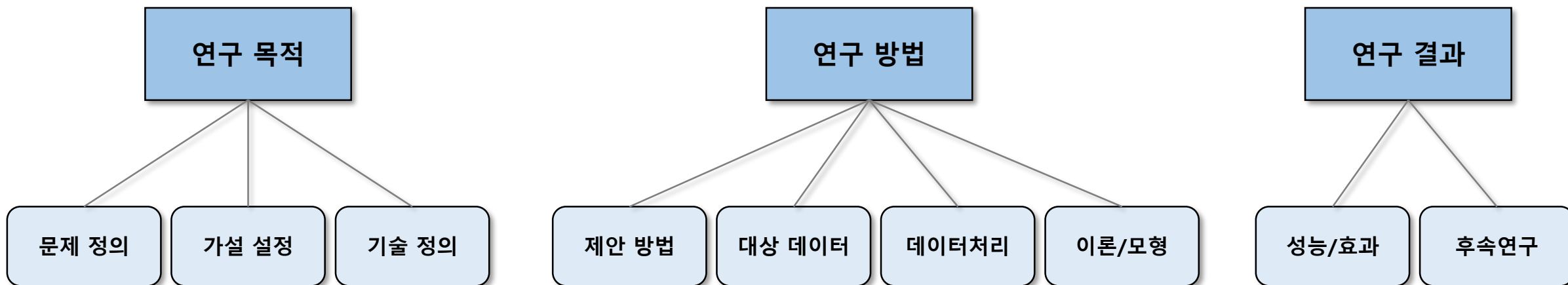
- 텍스트 분류 (Text Classification)

- 문장이나 문서를 사람이 정한 몇 가지 범주 중 어느 범주에 속하는지 분류하는 문제
  - 감정 분류, 뉴스 분류 등



# 문제 정의

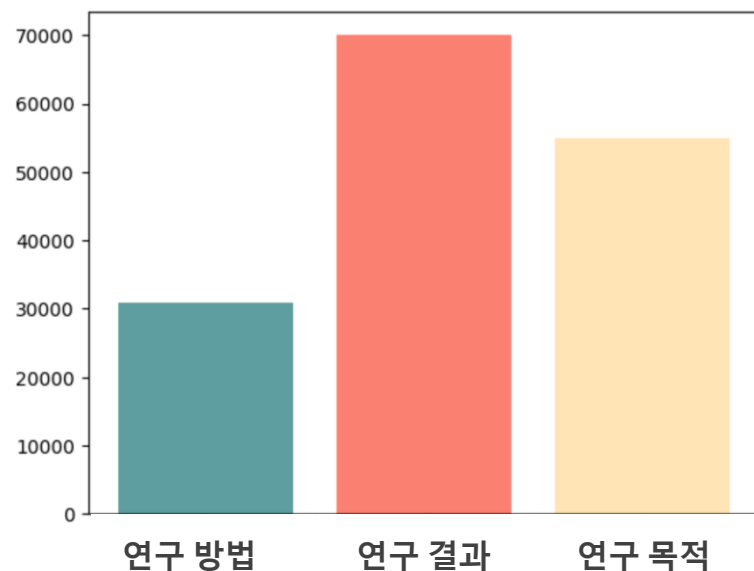
- 과학기술 논문 내 문장들의 수사학적 카테고리 예측
- 계층적 카테고리 구조를 고려한 모델의 필요성
  - 계층적 카테고리 구조의 경우, 카테고리간 불균형 문제 존재
  - 상위/하위 카테고리간 관계 뿐만 아니라 같은 계층 사이의 관계 학습 필요



# 활용 데이터

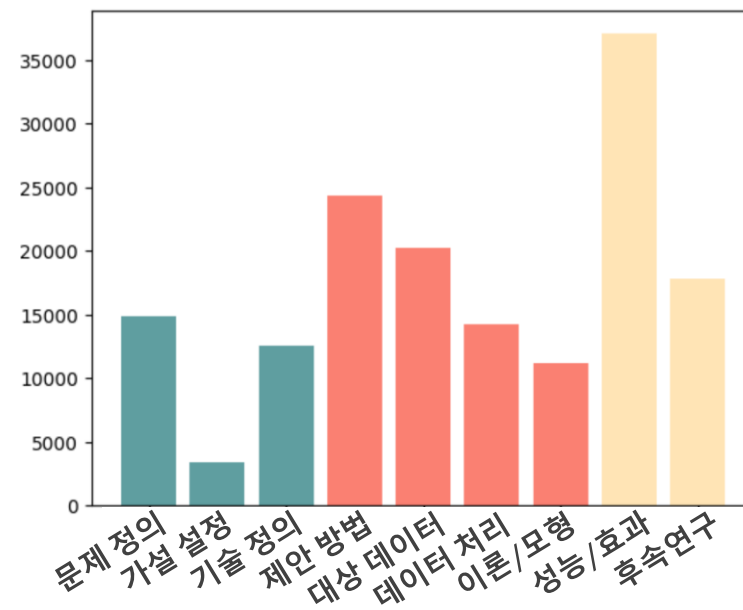
- 국내 논문 문장 의미 태깅 데이터셋
  - 별도의 전처리없이 원본 데이터 사용
    - 문장: 155,740개
    - 논문: 14,083건

## • 대분류



```
{
  "doc_id": "논문ID",
  "sentence": "문장 단위 텍스트",
  "tag": "문장 역할 태그",
  "keysentence": "태그별 대표 문장 여부(yes/no)"
}
```

## • 세부분류



# 활용 데이터

## ▪ 국내 논문 전문 텍스트 데이터셋 (481,578개)

- 문장 의미 태깅 데이터셋에 포함된 논문의 "문장 단위 텍스트"와 "본문 텍스트"를 대조하여,
- 논문 내 문장의 상대적 위치와 섹션명을 가져와 추가 실험에 사용하였다.

예시) 0.9781 [SEP] 결론 [SEP] 문장

논문 내 문장의 상대적 위치 =  $\frac{\text{문장이 속한 섹션의 인덱스}}{\text{전체 섹션 수}}$

```
{
  "doc_id": "논문ID",
  "sentence": "문장 단위 텍스트",
  "tag": "문장 역할 태그",
  "keysentence": "태그별 대표 문장 여부(yes/no)"
}
```

<국내 논문 문장 의미 태깅 데이터셋>

```
{
  "doc_id": "논문ID",
  "title": { "ko": "국문 제목", "en": "영문 제목" },
  "authors": "저자",
  "journal": { "ko": "국문 학술지/학술대회 제목", "en": "영문 학술지/학술대회 제목" },
  "year": "발행연도",
  "abstract": { "ko": "국문 초록", "en": "영문 초록" },
  "keywords": { "ko": "국문 키워드", "en": "영문 키워드" },
  "body_text": [
    {
      "section": "장, 절, 단락",
      "text": [ "해당 section의 텍스트" ]
    }
  ]
}
```

<국내 논문 전문 텍스트데이터셋>

## 제안 모델

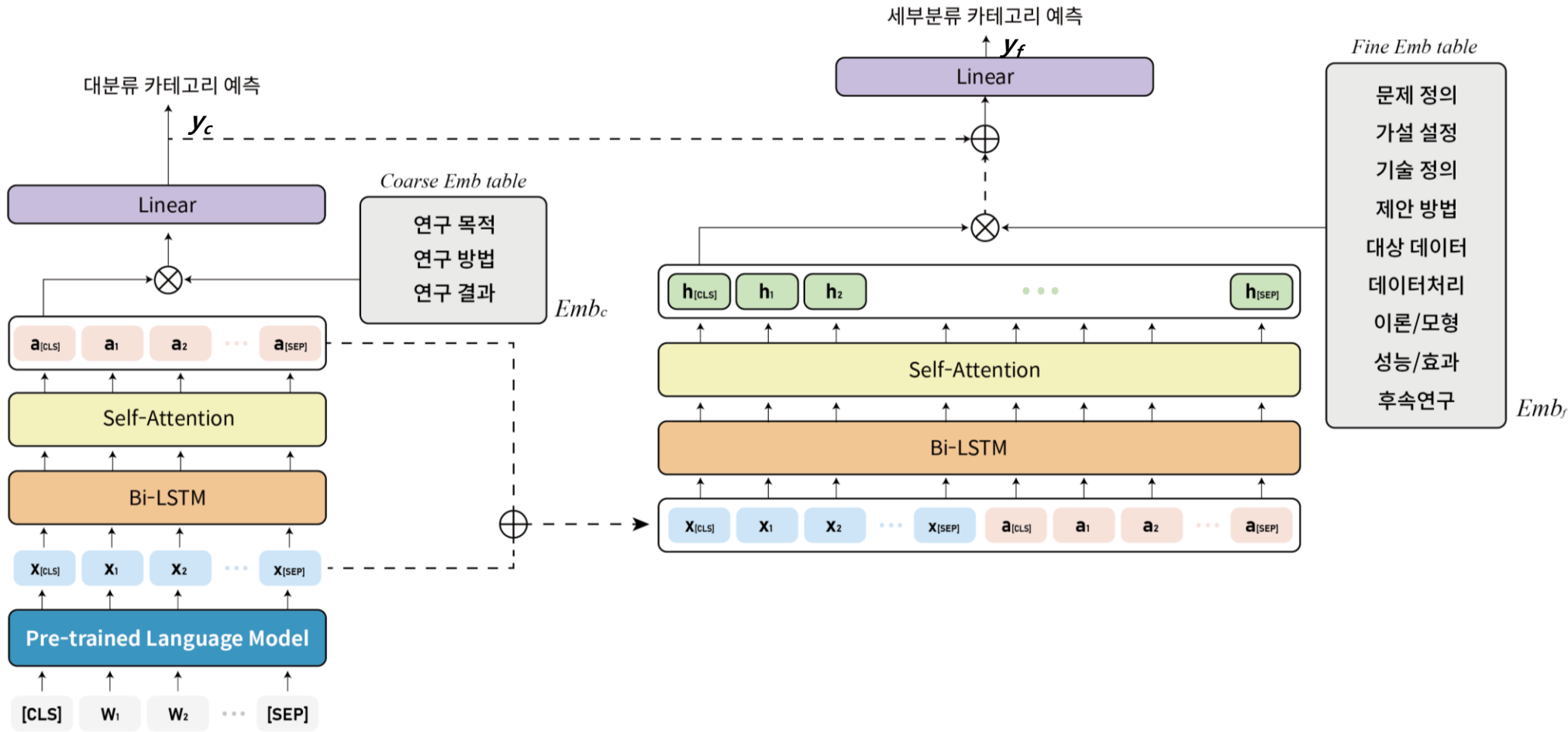
- 계층적 loss, representation 및 레이블 임베딩을 활용한 수사학적 논문 문장 분류 모델

*Hierarchical classification with hierarchical loss, representation and trainable label embedding*

- ✓ 계층적 손실함수를 통한 카테고리 예측 및 계층 구조 학습
- ✓ 계층적 representation을 통한 대분류와 세부분류 사이 정보 공유
- ✓ KorSciBERT와 Bi-LSTM을 활용한 레이블 임베딩
- ✓ 논문 전문 데이터셋의 섹션명과 논문 내 문장의 상대적 위치 활용

# 제안 모델

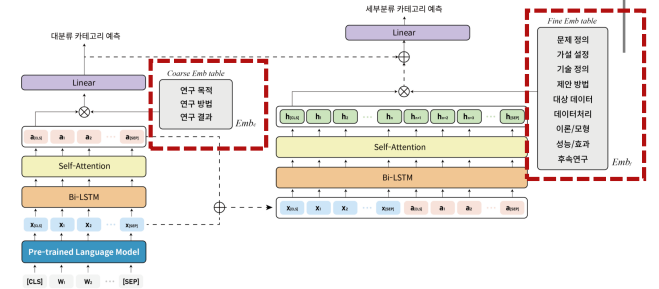
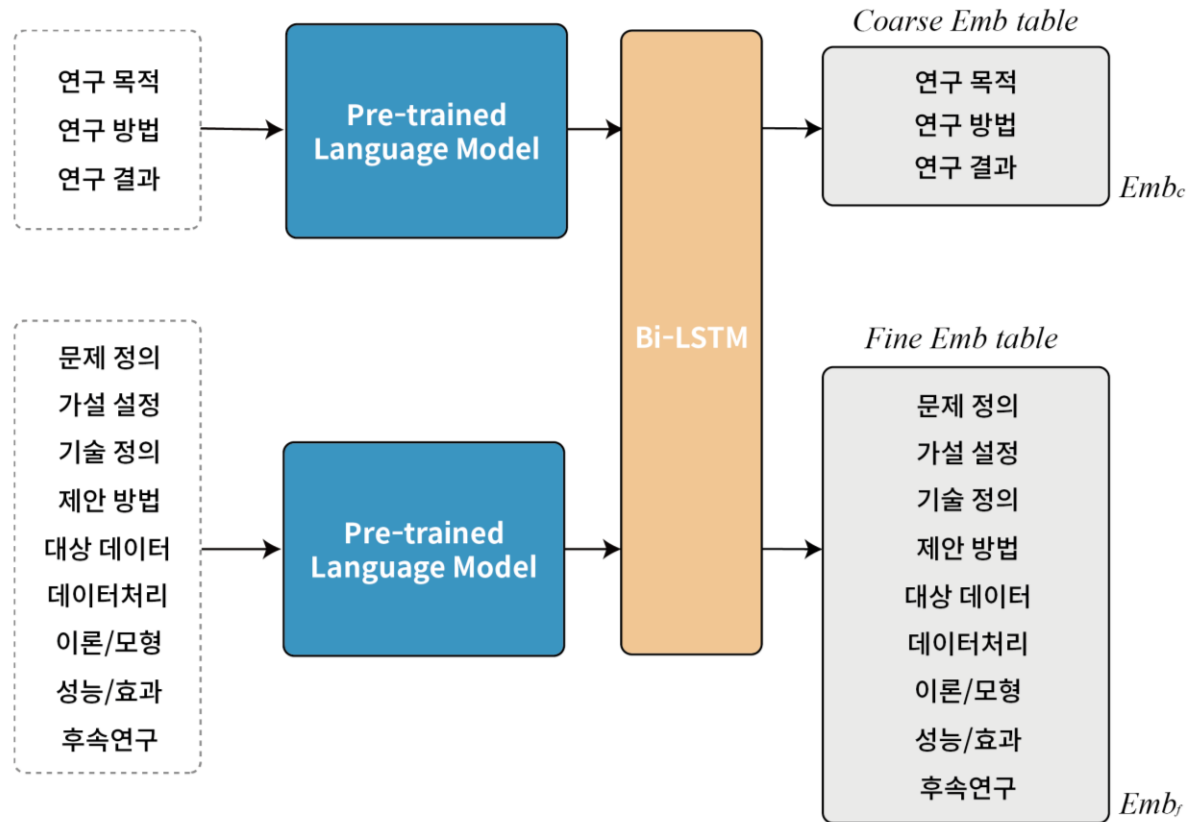
▪ 전체 모델 구조





# 제안 모델

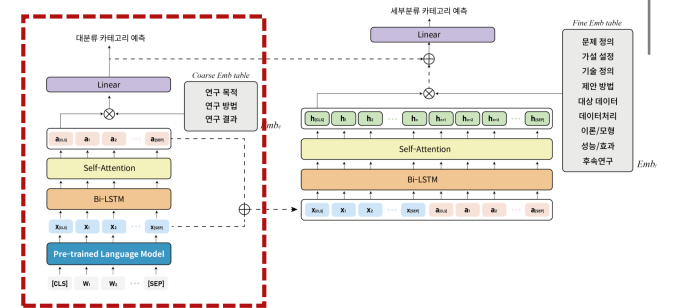
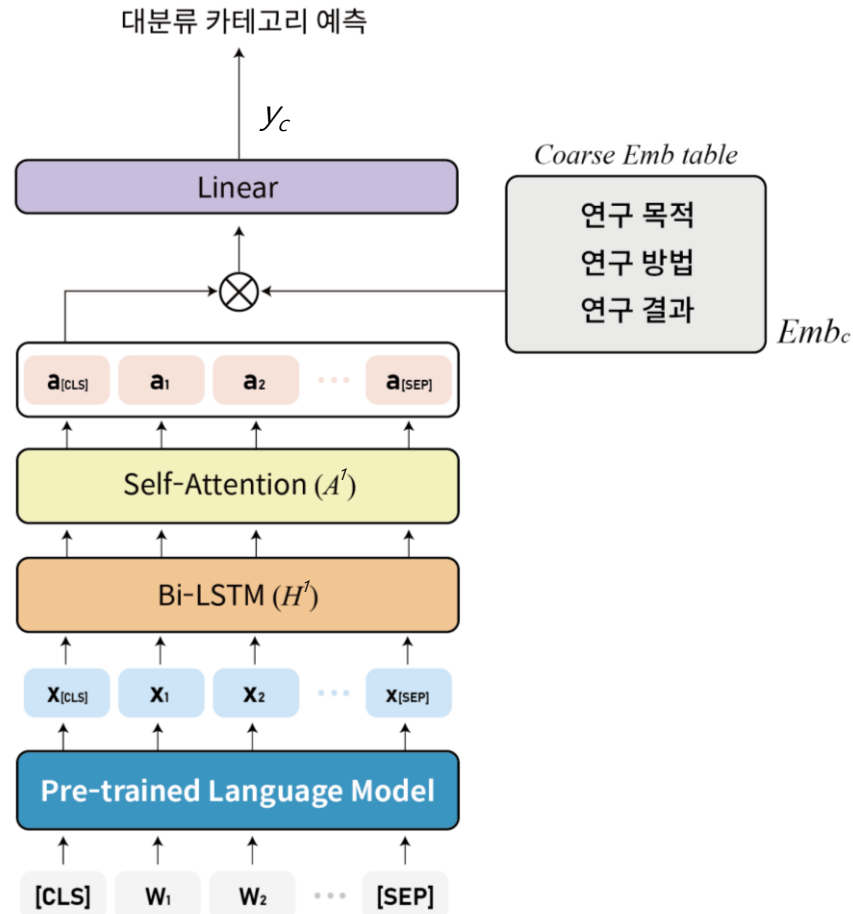
- 사전 학습 모델을 활용한 레이블 임베딩 테이블



- 사전 학습 언어모델(KorSciBERT)을 사용하여 레이블 임베딩 테이블 초기화
- 학습과 함께 파라미터 업데이트

# 제안 모델

## ■ 대분류 예측



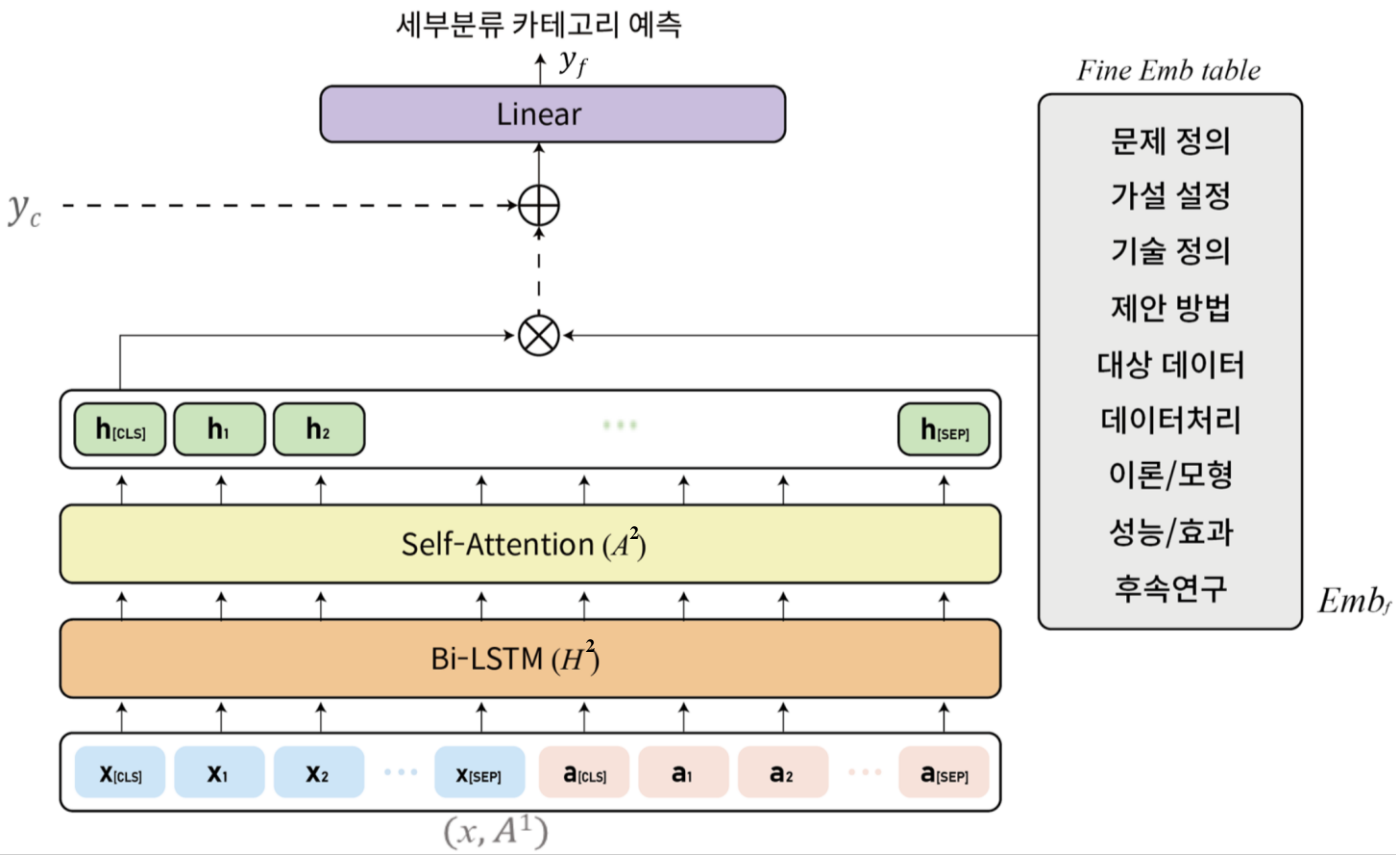
- KorSciBERT를 통한 논문 문장 학습
- Bi-LSTM과 Self-attention을 통한 문장 내 중요 정보 포착

$$A^1 = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_h}}\right) \cdot V$$

$$y_c = \text{Linear}(A^1 \cdot Emb_c)$$

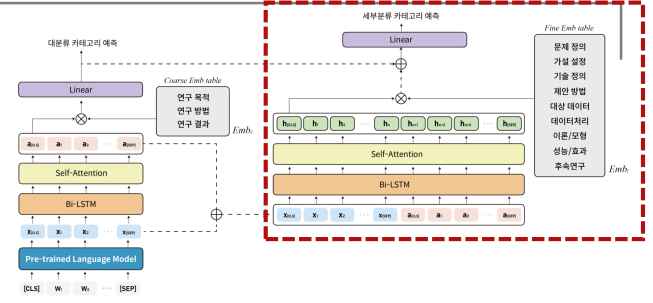
# 제안 모델

- 세부 분류 예측



- 계층적 representation 통한 information sharing
- 대분류 예측 정보를 세부분류 예측에 활용

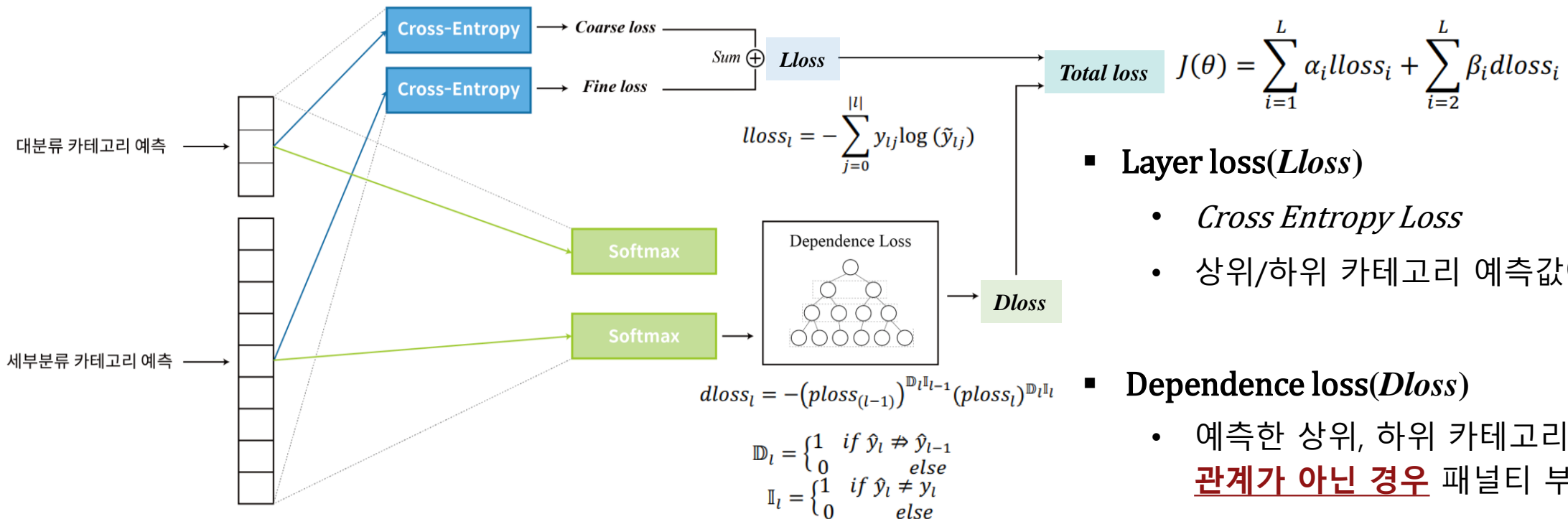
$$A^2 = softmax(\frac{Q \cdot K^T}{\sqrt{d_h}}) \cdot V$$
$$y_f = Linear(A^2 \cdot Emb_f \cdot y_c)$$



# 제안 모델

## 계층적 손실 함수

- 계층적 손실 함수를 통해 대분류 및 세부분류 예측 성능 향상
- 상위 카테고리 and 하위 카테고리 사이의 구조 학습



# 실험

- 데이터셋

- 국내 논문 문장 의미 태깅 데이터셋
- 총 155,739건의 데이터를 9 : 0.5 : 0.5 비율로 나누어 교차 검증 진행

|    |         |
|----|---------|
| 전체 | 155,739 |
| 훈련 | 140,166 |
| 검증 | 7,787   |
| 평가 | 7,787   |



Train : Dev : Test = 9 : 0.5 : 0.5

# 성능 평가

## ▪ 평가 지표

$$Macro\ F1 = \frac{1}{n} \sum_{i=1}^n \frac{(2 \times precision_i \times recall_i)}{precision_i + recall_i}$$

$$Micro\ F1 = \frac{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n TN_i + \sum_{i=1}^n FP_i + \sum_{i=1}^n FN_i}$$

|          | Positive       | Negative       |
|----------|----------------|----------------|
| Positive | True Positive  | False Positive |
| Negative | False Negative | True Negative  |

# 성능 평가

- Metric: F1-score
- Baseline: 전년도 국내 논문 의미 태깅 수상작

| PLM                              | 대분류          |              | 세부분류         |              |
|----------------------------------|--------------|--------------|--------------|--------------|
|                                  | macro F1     | micro F1     | macro F1     | micro F1     |
| Baseline<br>KorSciBERT           | 96.02        | 95.51        | 89.95        | 89.81        |
| Baseline<br>KLUE roBERTa-base    | 95.64        | 96.43        | 89.77        | 89.58        |
| <b>Our<br/>KorSciBERT</b>        | <b>95.56</b> | <b>96.10</b> | <b>90.51</b> | <b>90.30</b> |
| <b>Our<br/>KLUE roBERTa-base</b> | 95.25        | 95.89        | 90.05        | 89.94        |

# 성능 평가

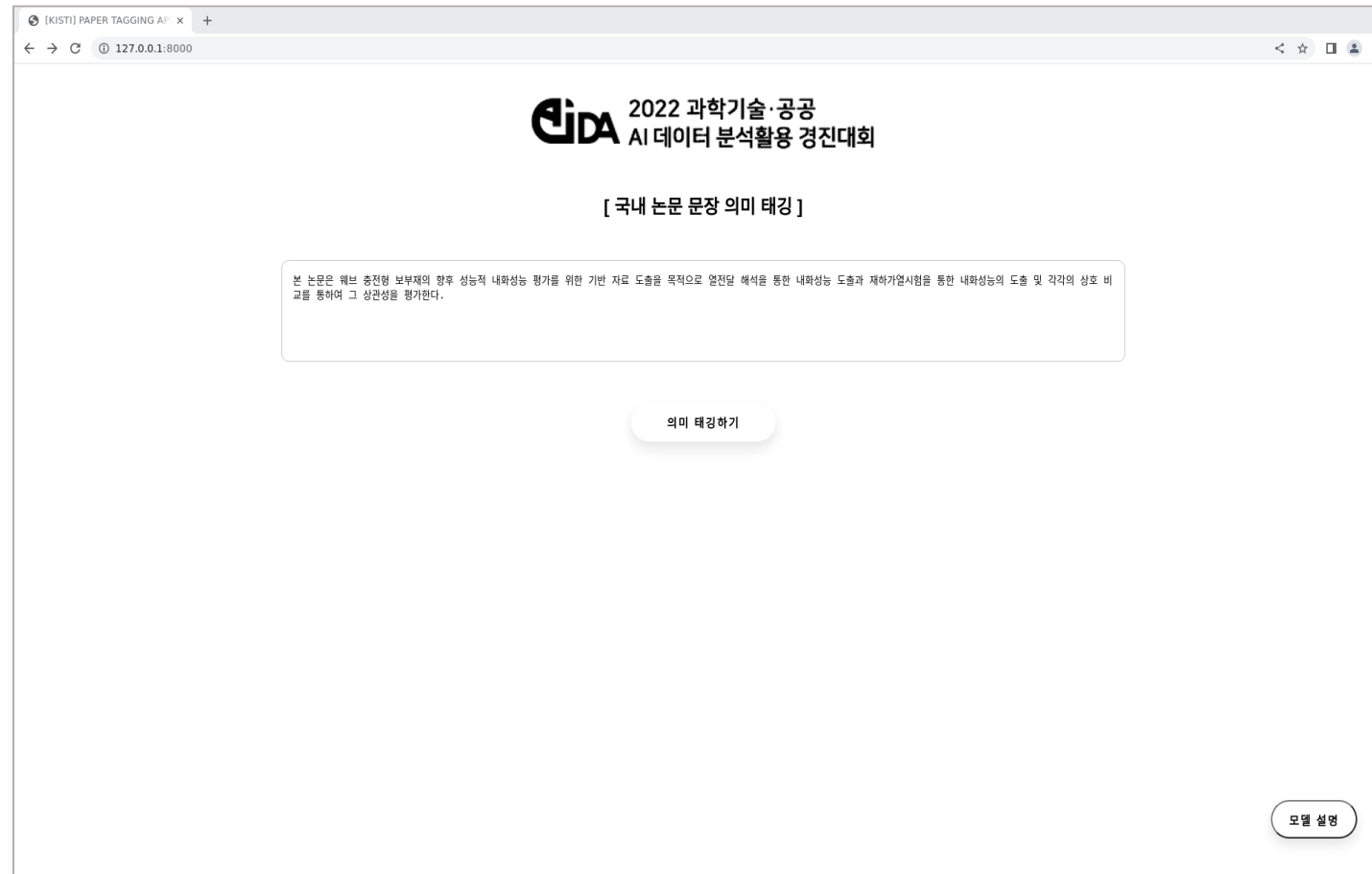
- 섹션명과 논문 내 문장의 상대적 위치 활용 실험

| PLM                               | 대분류          |              | 세부분류         |              |
|-----------------------------------|--------------|--------------|--------------|--------------|
|                                   | macro F1     | micro F1     | macro F1     | micro F1     |
| Our KorSciBERT                    | 95.56        | 96.10        | 90.51        | 90.30        |
| KorSciBERT with section           | 96.56        | 97.02        | 91.21        | 90.93        |
| KorSciBERT with section, position | <b>96.61</b> | <b>97.04</b> | <b>91.59</b> | <b>91.31</b> |



# 웹페이지 구현

- 토큰별 어텐션 스코어 시각화를 통한 문장 분류의 근거 제공



# 웹페이지 구현

- 토큰별 어텐션 스코어 시각화를 통한 문장 분류의 근거 제공

The screenshot shows a web browser window with the URL `127.0.0.1:8000/predict-by-sent`. The page title is "2022 과학기술·공공 AI 데이터 분석활용 경진대회" and the subtitle is "[ 국내 논문 문장 의미 태깅 ]". There is a text input field with the placeholder "의미 태깅할 문장을 입력해주세요." and a button labeled "의미 태깅하기". Below the input field is a section titled "입력 문장" containing a paragraph of text. To the right of the input field, a yellow box labeled "모델 예측 결과" contains the text "대분류 : 연구 목적" and "세부분류 : 문제 정의". Below the input field, a purple box labeled "목적으로" points to the text "목적으로" in the input text. At the bottom right, there is a button labeled "모델 설명".

**모델 예측 결과**

대분류 : 연구 목적  
세부분류 : 문제 정의

**목적으로**

**입력 문장**

본 논문은 웹브라우저 상의 상호작용 성능 평가를 위한 기반 자료 도출을 목적으로 열전달 해석을 통한 내화 성능 도출과 재하가열시험을 통한 내화 성능의 도출 및 각각의 상호 비교를 통하여 그 상관성을 평가한다.

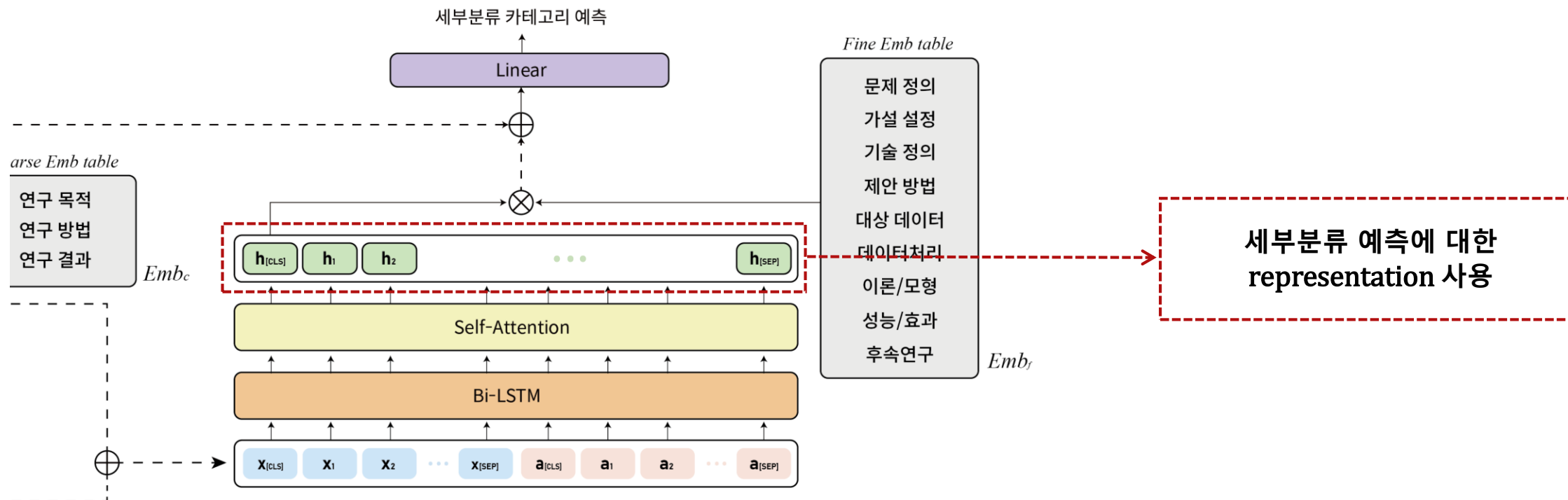
**입력 문장의 어텐션 시각화**

[CLS] 본 논문은 웹브라우저 상의 상호작용 성능 평가를 위한 기반 자료 도출을 목적으로 열전달 해석을 통한 내화 성능 도출과 재하가열시험을 통한 내화 성능의 도출 및 각각의 상호 비교를 통하여 그 상관성을 평가한다. [SEP]

**모델 설명**

# 웹페이지 구현

- 토큰별 어텐션 스코어 시각화를 통한 문장 분류의 근거 제공



- 대분류에 대한 정보와 세부분류에 대한 정보를 모두 담고 있는 representation

# 결론

## ■ 대분류/세부분류의 계층적 구조를 고려한 모델

- 문장에 대한 representation을 구하고 대분류 예측 정보를 세부분류 예측에 사용
- 계층적 손실 함수를 통해 카테고리의 계층형 구조 학습

## ■ KorSciBERT와 Bi-LSTM을 활용한 레이블 임베딩

- KorSciBERT를 사용하여 레이블 임베딩 테이블 초기화
- Bi-LSTM을 통해 임베딩 테이블 파라미터 업데이트

## ■ 문장의 위치 정보(섹션의 상대적 위치 + 섹션명)를 활용

- 논문 전문 데이터셋에서 해당하는 문장에 대한 섹션명을 가져옴
- 해당 섹션의 상대적인 위치를 실숫값으로 나타낸 후 사용함
- 문장 분류 성능 향상에 도움을 줌

## 활용계획 및 기대효과

- 논문 정보의 효율적 검색, 주요 문장 추출, 논문 요약, 트렌드 분석, 저자의 의도 파악 등 다양한 태스크에서 활용 가능
- 사용자는 웹에서 제공되는 어텐션 시각화를 통해 **분류의 근거**를 확인할 수 있음
- **계층적 카테고리**가 주를 이루는 e-commerce와 같은 여러 영역에서 범용적으로 활용 가능

## 향후 연구

### ▪ 기계적 논문 생성 연구 분야로 확장 가능

- 논문 문장 의미 태깅을 통해 생성되는 텍스트가 특정 주제를 벗어나지 않도록 제어함으로써 생성 성능을 향상시킬 수 있다.
- 문맥 기반 의미 태깅 모델은 실제 논문의 구성을 잘 반영하는 논문 생성 연구 분야에 활용 가능할 것으로 기대
  - 논문을 제한적인 구성을 따르므로 논문의 각 문장을 태깅하면 의미 태그들의 등장 패턴 파악 가능
  - 이를 바탕으로 다음에 등장할 문장의 의미 태그 예측이 가능하다면, 예측한 태그의 범주 내에서 다음 문장을 생성하도록 제한을 가할 수 있을 것으로 기대

# 시연

# Q & A



**감사합니다.**

## 참고자료

- Gao, Dehong, et al. "Deep hierarchical classification for category prediction in E-commerce system." *arXiv preprint arXiv:2005.06692* (2020).
- Miyazaki, Taro, et al. "Label embedding using hierarchical structure of labels for twitter classification." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019.
- 김담린, 박성식, and 김학수. "계층적 다중 레이블 임베딩을 이용한 논문 문장 수사학적 분류 모델." 한국정보과학회 학술발표논문집 (2022): 401-403.
- 전년도 베이스라인
  - <https://aida.kisti.re.kr/data/8d0fd6f4-4bf9-47ae-bd71-7d41f01ad9a6/gallery/17>