

2024 한국컴퓨터종합학술대회

다중 레이블 분류를 위한 프롬프팅 고도화

Enhanced Prompting for Multi-Label Classification

2024.06.28

이정연, 신영우, 윤예진, 김태욱[†]

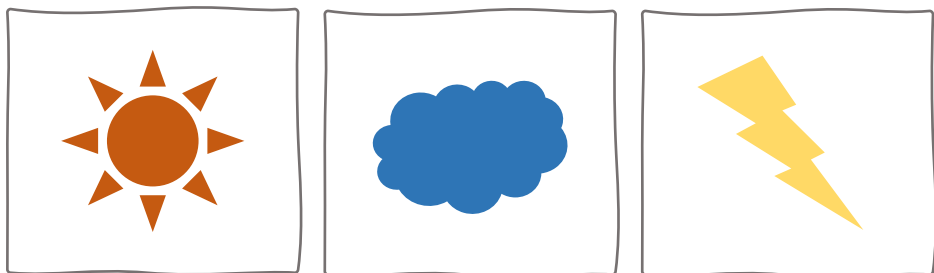
한양대학교 자연어처리 연구실

다중 레이블 분류 (Multi-Label Classification)

다중 클래스 분류 (Multi-Class Classification)

두 개 이상의 클래스가 있는 분류 작업에서, 하나의 입력에 하나의 레이블이 할당되는 작업

1 input single-label



[1 0 0]

[0 1 0]

[0 0 1]

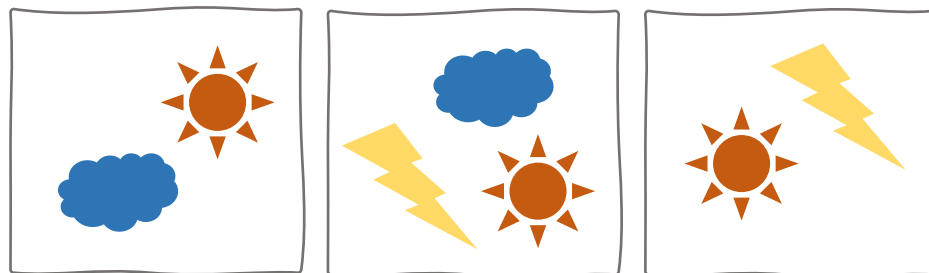
다중 레이블 분류 (Multi-Label Classification)

하나의 입력에 여러 개의 레이블(label)이

1 input

multi-label

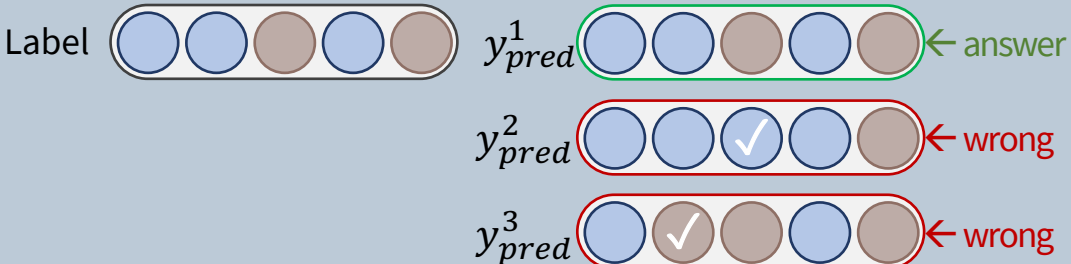
할당될 수 있는 분류 작업



[1 1 0]

[1 1 1]

[1 0 1]



다중 레이블 분류 작업의 어려움

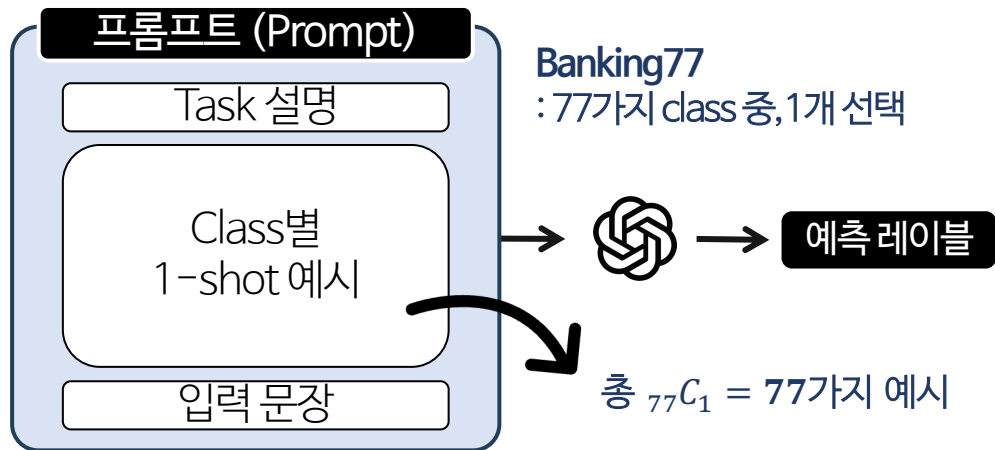
- 모델이 **모든 레이블**을 정확히 맞춰야 정답
- 하나의 레이블이라도 틀리면 오답으로 간주
→ 성능 평가에 보다 엄격한 기준이 적용됨.

상황 내 학습에서의 다중 레이블 분류

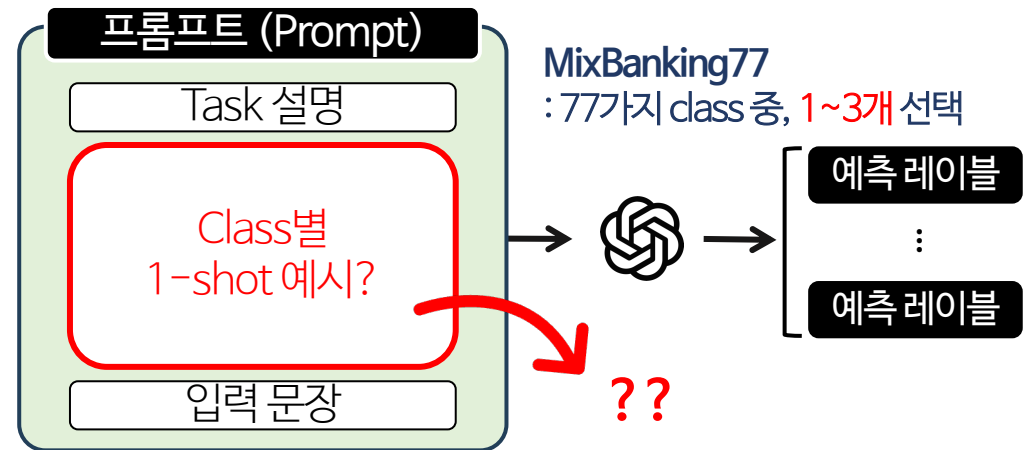
상황 내 학습에서의 다중 레이블 분류 작업의 **어려움** : Banking77과 MixBanking77에 대한 비교

분류 문제에서 1-shot의 상황 내 학습의 경우, 프롬프트에 각 클래스 별 1개의 예시가 포함

다중 클래스 분류 (Multi-Class Classification)



다중 레이블 분류 (Multi-Label Classification)



대형 언어 모델 (Large Language Models)의 **상황 내 학습 (In-context Learning)**

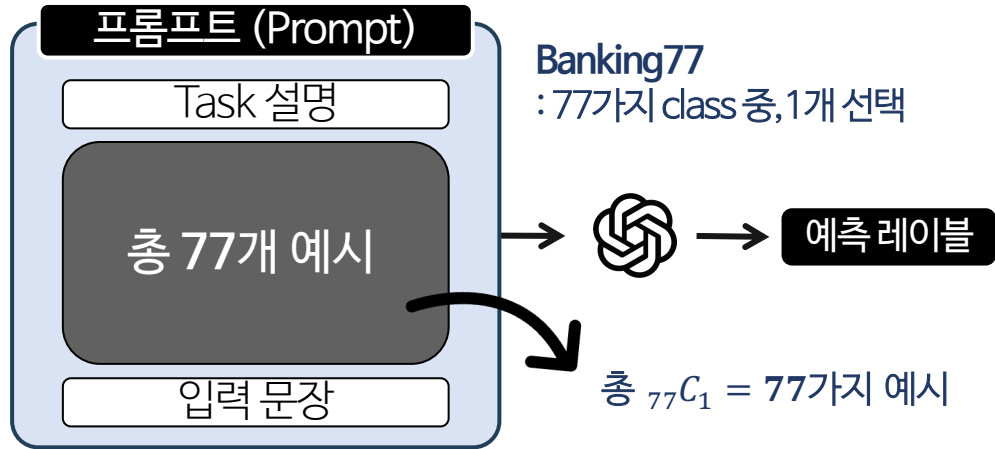
- 별도의 미세 조정 없이, 모델이 주어진 상황(context) 내 정보를 활용하여 작업을 수행하도록 하는 방법론
- 모델은 특정 작업 수행을 위해 주어진 예시(demonstration)에서 패턴을 스스로 유추

상황 내 학습에서의 다중 레이블 분류

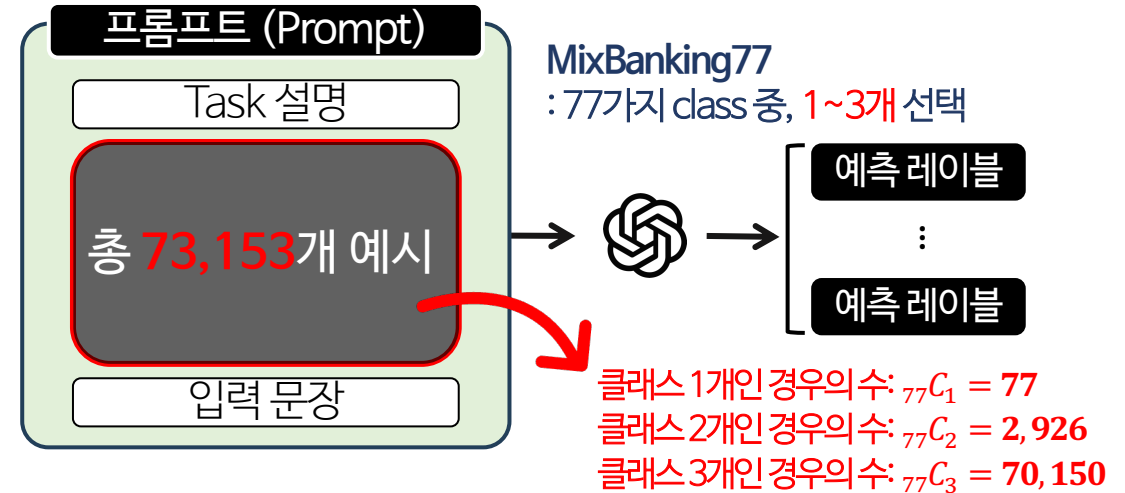
상황 내 학습에서의 다중 레이블 분류 작업의 어려움 : Banking77과 MixBanking77에 대한 비교

분류 문제에서 1-shot의 상황 내 학습의 경우, 프롬프트에 각 클래스 별 1개의 예시가 포함

다중 클래스 분류 (Multi-Class Classification)



다중 레이블 분류 (Multi-Label Classification)



대형 언어 모델 (Large Language Models)의 상황 내 학습 (In-context Learning)

- 별도의 미세 조정 없이, 모델이 주어진 상황(context) 내 정보를 활용하여 작업을 수행하도록 하는 방법론
- 모델은 특정 작업 수행을 위해 주어진 예시(demonstration)에서 패턴을 스스로 유추

기존 연구 한계

Fine-tuning 방식 대비 **크게 낮은** 상황 내 학습 성능

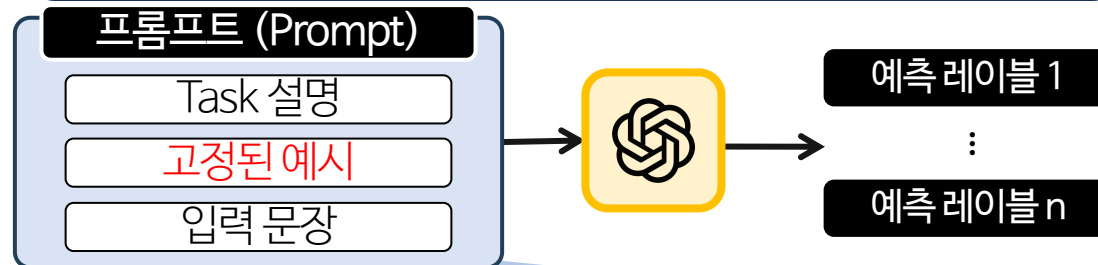
(Dataset: MixATIS)

Method		Score
In-Context Learning	(Yin et al., 2024)	54.10
	BlendX (Yoon et al., 2024)	40.30
GL-GIN (Qin et al., 2021)		76.30
Fine-tuning	SDJN (Chen et al., 2022)	77.10
	TFMN (Chen et al., 2022)	79.80

(Dataset: MixSNIPS)

Method		Score
In-Context Learning	(Yin et al., 2024)	83.90
	BlendX (Yoon et al., 2024)	81.68
GL-GIN (Qin et al., 2021)		95.60
Fine-tuning	SDJN (Chen et al., 2022)	96.50
	TFMN (Chen et al., 2022)	97.70

기존 상황 내 학습 연구에서 프롬프팅 예시



You are an Intent Detection Model on single utterance.

Task 설명

[Task Definition] Detect single or more intent(s) of each utterance, but you can only classify UP TO 3 most plausible intents on 1 utterance.

[Intents] atis_airport, atis_ground_service, atis_abbreviation, atis_city, atis_aircraft, atis_ground_fare, atis_flight, ...

[Answer format] If more than one, concatenate with '#', such as {Intent}#{Intent}.
e.g. atis_ground_fare#atis_distance

[Example 1]

[Utterance] does delta aircraft fly dc10
[Answer] atis_aircraft

고정된 예시

[Example 2]

[Utterance] which airline has more business class flights than any other airline and what city is the airport mco in
[Answer] atis_airline#atis_city

[Example 3]

[Utterance] what does the fare code qx mean, what is the distance between pittsburgh airport and downtown pittsburgh and what is restriction ap80
[Answer] atis_abbreviation#atis_distance#atis_restriction

[Query] Detect a single or up to 3 intent(s) on this following utterance. : utt

입력 문장

본 연구 요약

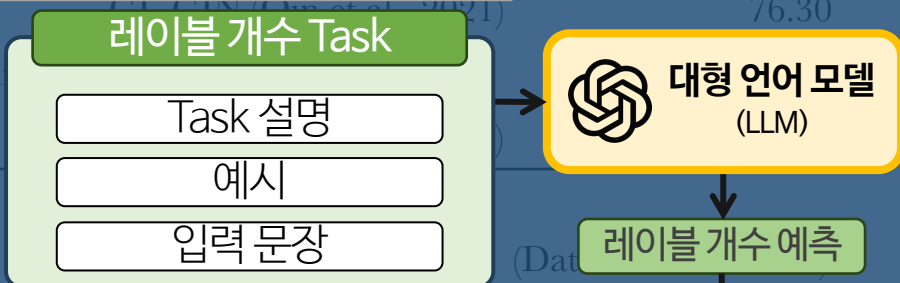
Fine-tuning 방식 대비 크게 낮은 상황 내 학습 성능 비교기준

(Dataset: MixATIS)

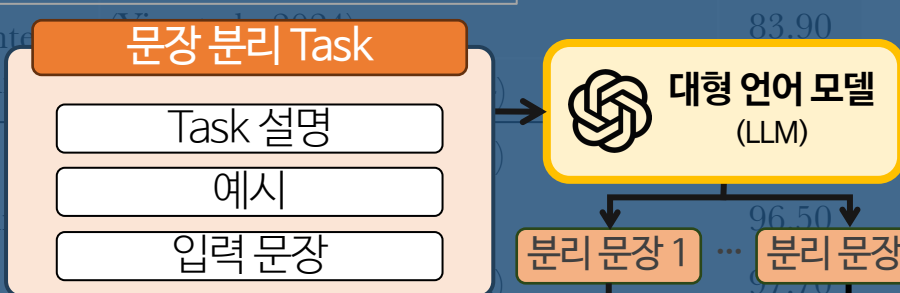
Method	Score
In-Context Learning (Yin et al., 2024)	54.10
연쇄 프롬프팅 1 개수 예측 기반	40.30
연쇄 프롬프팅 2 분리문장 기반	76.30

본 연구 제안

연쇄 프롬프팅 1 개수 예측 기반



연쇄 프롬프팅 2 분리문장 기반



단일 프롬프팅 상황 내 학습 연구에서 프롬프팅 예시

다중 레이블 분류 Task

- Task 설명
- 고정된 예시
- 입력 문장



예측 레이블 1

⋮

예측 레이블 n

다중 레이블 분류 Task

- Task 설명
- 유사도 기반 예시
- 입력 문장 **개수 예측**



예측 레이블 1

⋮

예측 레이블 n

단일 레이블 분류 Task

- Task 설명
- 유사도 기반 예시
- 분리 문장 1



예측 레이블 1

⋮

예측 레이블 n

방법론 1. 프롬프트 역할 구분

• ChatGPT 프롬프트 내 역할 구분

- 시스템(system): ChatGPT의 행동을 지정
- 사용자(user): 입력 문장(쿼리) 또는 예시 문장
- 보조자(assistant): 예시 문장의 응답(레이블)

```
{  
  "role": "system", "content": "You are an Intent Detection Model ..."  
  "role": "user", "content": "is there limo service at pittsburgh airport ..."  
  "role": "assistant", "content": "atis_ground_service#atis_meal"  
}
```

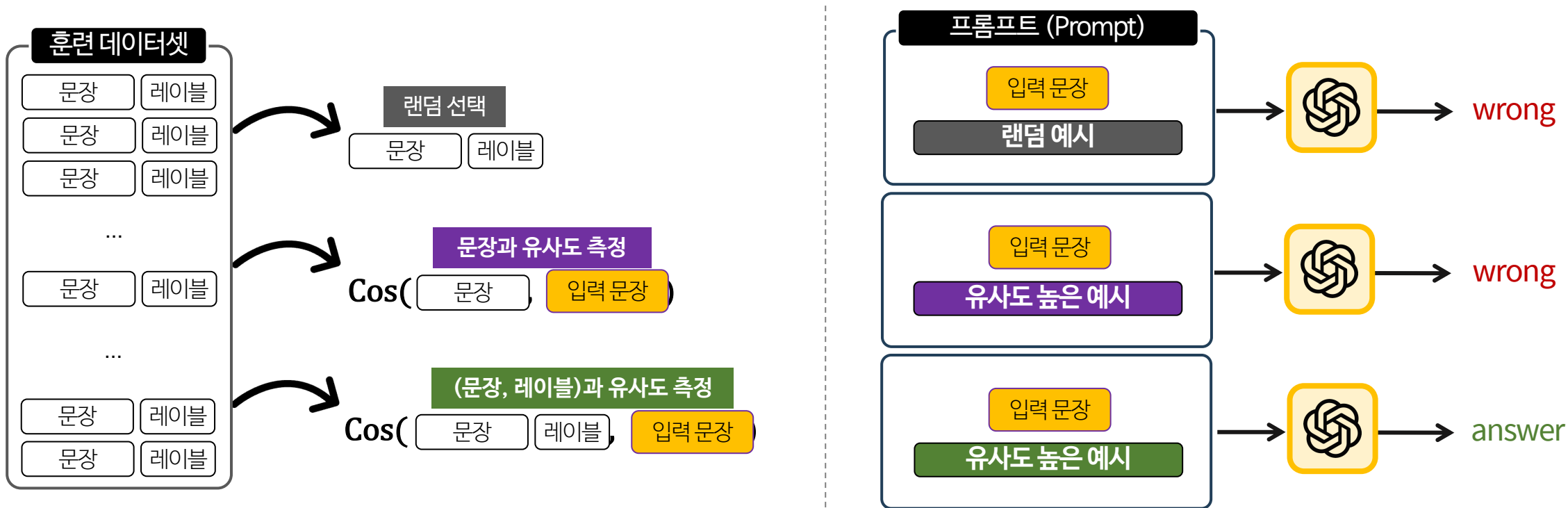
* 입력 문장에 따라 각각의 예시 문장 및 레이블 제공

OpenAI에서 공개한 ChatGPT API 사용 가이드라인을 따라
프롬프트 내 역할을 부여함으로써 맥락을 잘 이해하여 작업을 효과적으로 수행

방법론 2. 유사한 예시 선택

• 프롬프트 내 예시 선택 방식

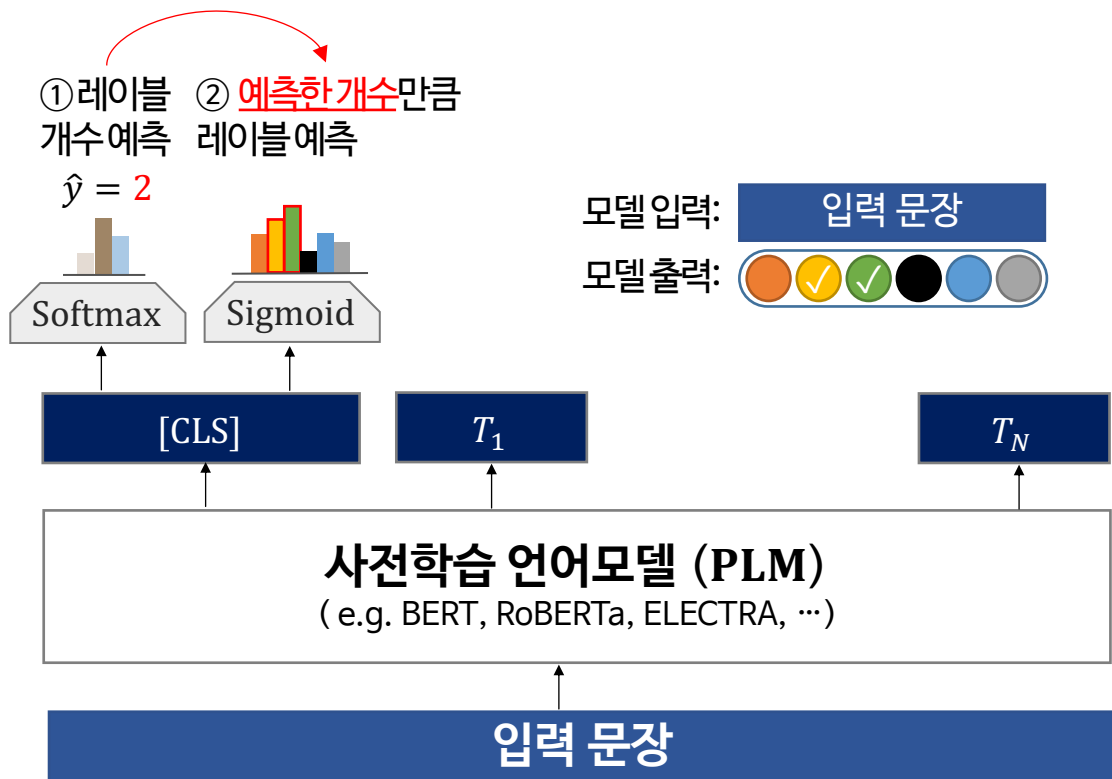
- 입력 문장에 대해 훈련 데이터셋의 (문장, 레이블) 쌍 중 코사인 유사도가 높은 k 개의 예시를 선택



입력 문장과 유사한 예시 검색 시 **문장**과 **레이블** 정보를 모두 활용함.

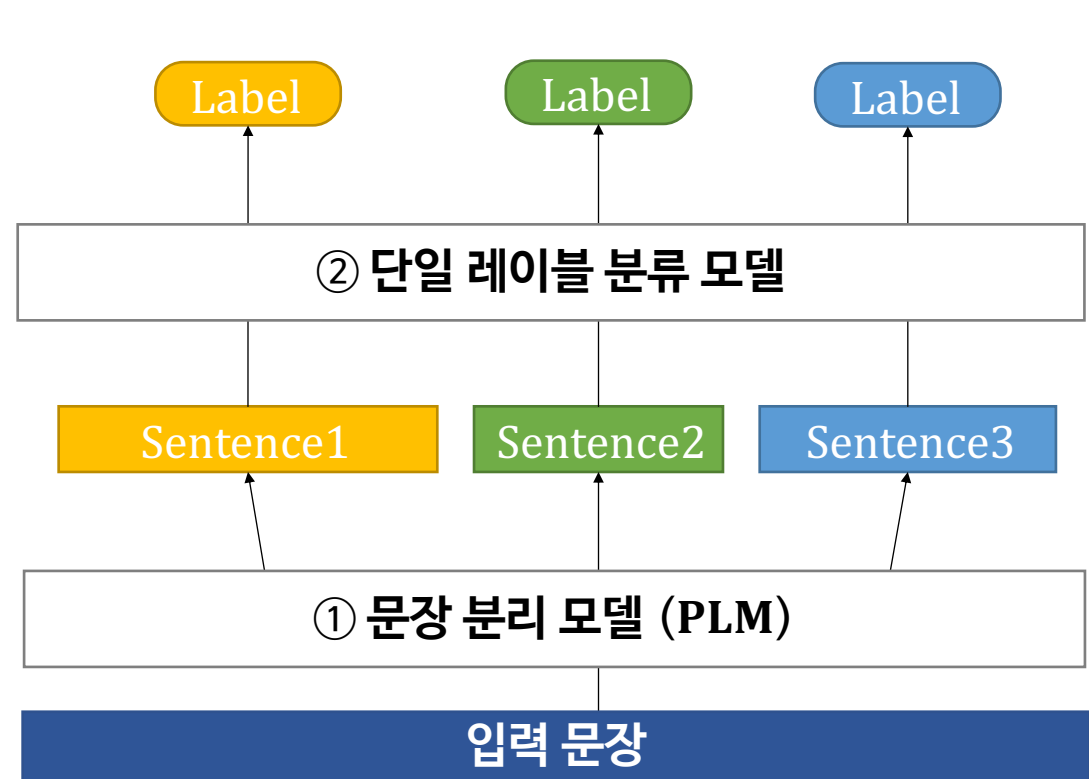
방법론 3. 연쇄 프롬프팅 기법 - fine-tuning 방법론에서 착안

TFMN (Chen et al., 2022)



입력 문장에 담긴 레이블 개수를 예측
→ 해당 개수만큼 확률 값이 높은 레이블을 최종 예측

DialogUSR (Meng et al., 2022)



입력된 다중 레이블 문장을 여러 개의 단일 레이블 문장으로 분리
→ 분리된 각 문장에 단일 레이블 분류를 진행하는 방법론 처음 제안


방법론 3. 연쇄 프롬프팅 기법

1. 레이블 개수 예측 기반 연쇄 프롬프팅

2. 문장 분리 기반 연쇄 프롬프팅


[입력 문장] what is meal code sb and also list airports in new york

① 레이블 개수 예측 ▶ 입력 문장을 여러 개의 단일 문장들로 분리하시오.

 대형 언어 모델 (LLM)


2개 입니다.

② 예측 개수 기반 다중 레이블 분류 ▶ 입력 문장에 담긴 **2개** 레이블이 무엇인지 예측하시오.

 대형 언어 모델 (LLM)


atis_abbreviation atis_airport

① 문장 분리 ▶ 입력 문장을 여러 개의 단일 문장들로 분리하시오.

 대형 언어 모델 (LLM)

list airports in new york
what is meal code sb

② 분리된 문장의 레이블 분류 ▶ 주어진 문장에 대해 레이블을 1개만 예측하시오.

 대형 언어 모델 (LLM)

atis_abbreviation atis_airport

해당 문장의 레이블은 atis_abbreviation, atis_airport 입니다.

실험 결과

• Main Result Table: 각 데이터셋에 대한 다중 레이블 분류 결과

- 제안한 3가지 프롬프팅 방법론을 적용한 실험 결과
- 기존 Fine-tuning 방식의 베이스라인 성능을 뛰어 넘을 뿐 아니라, 관련 선행 연구의 정확도 또한 큰 폭으로 개선

역할 구분	프롬프트 구성		예시 선택	MixATIS	MixSNIPS	MixBanking77	MixCLINC150
Fine-tuning				77.98	95.68	76.61	85.88
In-Context Learning				40.30	81.68	30.90	49.22
X	단일 프롬프트		랜덤**	36.70	77.56	23.72	45.55
O			랜덤	67.90	82.20	36.64	57.26
			유사도 기반	71.80	82.32	42.32	67.80
	연쇄 프롬프트	개수 예측 기반	랜덤	77.40	79.04	26.73	56.76
		유사도 기반	78.50	85.36	51.84	74.12	
	문장 분리 기반	랜덤	68.00	82.68	61.80	61.77	
		유사도 기반	80.40 ✓	87.40	66.23	78.29	

[표 1 각 데이터셋에 대한 다중 레이블 분류 결과 (metric: 정확도)]

실험 결과

• 프롬프트 내 역할 구분의 효과

- 역할 구분 없이 시스템(system)에 모든 프롬프트를 입력하였을 때 가장 낮은 정확도를 보임

역할 구분	프롬프트 구성		예시 선택	MixATIS	MixSNIPS	MixBanking77	MixCLINC150
	Fine-tuning			77.98	95.68	76.61	85.88
	In-Context Learning			40.30	81.68	30.90	49.22
X	단일 프롬프트		랜덤**	36.70	77.56	23.72	45.55
O			랜덤	67.90	82.20	36.64	57.26
			유사도 기반	71.80	82.32	42.32	67.80
	연쇄 프롬프트	개수 예측 기반	랜덤	77.40	79.04	26.73	56.76
		유사도 기반	78.50	85.36	51.84	74.12	
	문장 분리 기반	랜덤	68.00	82.68	61.80	61.77	
		유사도 기반	80.40	87.40	66.23	78.29	

[표 1 각 데이터셋에 대한 다중 레이블 분류 결과 (metric: 정확도)]

실험 결과

• 예시 선택 방식의 효과

- (1) 동적 예시 제공의 중요성: 매 입력 문장마다 랜덤으로 선정된 고정 예시를 제공할 때보다, 매번 다른 랜덤 예시를 제공하는 경우 성능 소폭 향상

역할 구분	프롬프트 구성		예시 선택	MixATIS	MixSNIPS	MixBanking77	MixCLINC150
Fine-tuning				77.98	95.68	76.61	85.88
In-Context Learning				40.30	81.68	30.90	49.22
X	단일 프롬프트		랜덤**	36.70	77.56	23.72	45.55
			랜덤	67.90	82.20	36.64	57.26
			유사도 기반	71.80	82.32	42.32	67.80
O	연쇄 프롬프트	개수 예측 기반	랜덤	77.40	79.04	26.73	56.76
			유사도 기반	78.50	85.36	51.84	74.12
		문장 분리 기반	랜덤	68.00	82.68	61.80	61.77
			유사도 기반	80.40	87.40	66.23	78.29

[표 1 각 데이터셋에 대한 다중 레이블 분류 결과 (metric: 정확도)]

실험 결과

• 예시 선택 방식의 효과

- (2) 유사도 기반 선택의 중요성: 매 입력 문장마다 유사한 예시를 제공하는 경우 정확도가 향상됨
 - 랜덤 선택 예시 결과에 비해 평균 7.6점의 정확도 향상

역할 구분	프롬프트 구성		예시 선택	MixATIS	MixSNIPS	MixBanking77	MixCLINC150
Fine-tuning				77.98	95.68	76.61	85.88
In-Context Learning				40.30	81.68	30.90	49.22
X	단일 프롬프트		랜덤**	36.70	77.56	23.72	45.55
			랜덤	67.90	82.20	36.64	57.26
O	단일 프롬프트		유사도 기반	71.80	82.32	42.32	67.80
			랜덤	77.40	79.04	26.73	56.76
	연쇄 프롬프트	개수 예측 기반	유사도 기반	78.50	85.36	51.84	74.12
			랜덤	68.00	82.68	61.80	61.77
	연쇄 프롬프트	문장 분리 기반	유사도 기반	80.40	87.40	66.23	78.29
			랜덤				

[표 1 각 데이터셋에 대한 다중 레이블 분류 결과 (metric: 정확도)]

실험 결과

• 연쇄 프롬프팅의 효과

- 유사도 기반 예시 선택 기법과 연쇄 프롬프팅을 함께 적용했을 때 가장 우수한 성능 확인

[표 2 연쇄 프롬프팅 세부 정확도]

MixX	ATIS	SNIPS	Banking77	CLINC150
개수 예측	94.80	93.20	84.08	91.20
문장 분리	94.90	77.40	68.82	89.43

역할 구분	프롬프트 구성		예시 선택	MixATIS	MixSNIPS	MixBanking77	MixCLINC150
Fine-tuning				77.98	95.68	76.61	85.88
In-Context Learning				40.30	81.68	30.90	49.22
X	단일 프롬프트		랜덤**	36.70	77.56	23.72	45.55
			랜덤	67.90	82.20	36.64	57.26
			유사도 기반	71.80	82.32	42.32	67.80
O	연쇄 프롬프트	개수 예측 기반	랜덤	77.40	79.04	26.73	56.76
			유사도 기반	78.50	85.36	51.84	74.12
	문장 분리 기반	랜덤	68.00	82.68	61.80	61.77	
		유사도 기반	80.40	87.40	66.23	78.29	

[표 1 각 데이터셋에 대한 다중 레이블 분류 결과 (metric: 정확도)]

Main Contributions

1

다중 레이블 분류를 상황 내 학습에 적용 시 발생하는 문제점 분석

- 수많은 레이블 중 정답인 다중 레이블 모두를 맞춰야 정답으로 인정됨
- 현실적으로 모든 레이블 조합 예시를 프롬프트 내에 담기 어려운 한계점

2

효과적인 프롬프팅 고도화 방안 제안

- 프롬프트 내 역할 구분
- 유사한 예시 선택
- 연쇄 프롬프팅 기법

3

제안 방법론의 우수성 확인

- 4가지 다중 의도 데이터셋 대상으로 실험
- MixATIS의 경우 Fine-tuning 방식의 성능 상회

Thank You

Jungyeon Lee

HYU NLP Lab.

Dept. of Artificial Intelligence
Hanyang University, South Korea

jungyune@hanyang.ac.kr