

과제 기술서: 대화 맥락 추론(외부 데이터 및 데이터 증강 사용 불가)

1. 과제 개요

‘대화 맥락 추론’ 과제는 입력으로 주어진 대화 내용을 바탕으로, 특정된 대상 발화로부터 다섯 가지의 추론문 유형인 ‘원인’, ‘후행 사건’, ‘전제 조건’, ‘내적 동기’, ‘감정 반응’ 중 하나에 상응하는 문장을 정확하게 추론하는 것을 목표로 한다. 즉 이 과제는 다지선다 형태로 진행되며, 각 대화에 대해 세 개의 추론 옵션이 제공되고 모델은 이 중 하나를 정답으로 선택해야 한다. 성공적인 추론은 대화의 구조와 흐름을 파악하는 능력뿐만 아니라, 사람들의 의사소통 방식에 대한 깊은 이해를 바탕으로 이루어진다. 따라서 과제를 통해 인공지능 모델이 대화의 문맥을 이해하는 능력과, 여러 추론 후보 중에서 가장 적합한 추론을 선택하는 능력을 평가할 수 있으며, 평가 결과를 바탕으로 향후 모델의 대화 맥락 이해 및 적합한 대화문 생성 능력을 강화할 수 있다.

2. 과제 정의

이 과제는 인공지능 모델로 하여금 대화문, 대상 발화, 그리고 추론문의 유형을 종합적으로 고려하여 가장 적합한 추론문을 선택하도록 하는 것을 목표로 한다. 즉 모델은 주어진 세 가지 추론문 후보 중 대상 발화의 추론 유형에 가장 적합한 하나를 선택해야 한다. 과제마다 하나의 적절한 추론문과 두 개의 부적절한 추론문이 제시되며, 모델이 참고하는 추론문 유형은 ‘원인’, ‘후행 사건’, ‘전제 조건’, ‘내적 동기’, ‘감정 반응’ 다섯 가지이다. 모델이 얼마나 정확하게 적절한 추론 문장을 선택하는지 정량적으로 측정하기 위한 지표로는 **정확도(Accuracy)**를 사용한다.

분류	내용	예시	비고
입력	대화	<pre>{ "speaker": 2, "utterance": "미스터트롯 콘서트 또 취소됐네", "utterance_id": "MDRW2100011357.1.1"}, { "speaker": 1, "utterance": "코로나 때문에 콘서트 다 취소돼서 그래", "utterance_id": "MDRW2100011357.1.2"}, { "speaker": 1, "utterance": "어쩔 수 없지 뭐", "utterance_id": "MDRW2100011357.1.3"}, { "speaker": 2, "utterance": "그래도 아직 두 군데 남아 있던데 ㅋㅋ", "utterance_id": "MDRW2100011357.1.4"}, { "speaker": 1, "utterance": "그런데 코로나 때문에 취소될 것 같아", "utterance_id": "MDRW2100011357.1.5"}, { "speaker": 2, "utterance": "취소되면 아쉽겠네 ㅠㅠ", "utterance_id": "MDRW2100011357.1.6"}, { "speaker": 1, "utterance": "서울은 한 주차라도 연기해서 코로나 끝나고 했으면 좋겠어", "utterance_id": "MDRW2100011357.1.7"}, { "speaker": 2, "utterance": "콘서트 운영자들도 계획이 있겠장~", "utterance_id": "MDRW2100011357.1.8"}, { "speaker": 1, "utterance": "코로나 앞에는 답이 없당 다른 콘서트들은 하던데", "utterance_id": "MDRW2100011357.1.9"}, { "speaker": 2, "utterance": "나훈아 콘서트도 말렸음 ㅋㅋ 뮤지컬만 하더랑", "utterance_id": "MDRW2100011357.1.10"}, { "speaker": 1, "utterance": "나훈아나 대형 가수급들은 못 하나 봐", "utterance_id": "MDRW2100011357.1.11"}, { "speaker": 1, "utterance": "소규모 공연만 가능한 듯", "utterance_id": "MDRW2100011357.1.12"}, { "speaker": 2, "utterance": "글쎄 빨리 코로나 끝나서 콘서트 했으면 좋겠네 ㅋ", "utterance_id": "MDRW2100011357.1.13"}, { "speaker": 1, "utterance": "미스터트롯 콘서트는 어른들 효도 콘서트나 다름없는데 할머니들 목 빠지시겠어", "utterance_id": "MDRW2100011357.1.14"}, { "speaker": 2, "utterance": "그러게~ 올해 어버이날 선물로 엄청 팔렸단데 ㅋ", "utterance_id": "MDRW2100011357.1.15"} }</pre>	문자열 리스트



		"MDRW2100011357.1.15"},	
	대상 발화 id	[MDRW2100011357.1.14]	문자열 리스트
	유형	원인	문자열
	추론 1	코로나로 콘서트가 연이어 취소되고 있다.	문자열
	추론 2	코로나로 콘서트가 연이어 개최되고 있다.	문자열
	추론 3	코로나로 뮤지컬이 연이어 취소되고 있다.	문자열
출력	정답	추론 1("inference_1")	문자열
평가		Accuracy	

표 1 대화 맥락 추론 과제 입력과 출력의 예

다섯 가지 추론문의 유형에 대한 정의는 다음과 같다.

유형	정의
원인(cause)	대상 발화의 사건을 유발하는 사건
후행 사건(subsequent event)	대상 발화 이후에 일어날 수 있는 사건
전제 조건(prerequisite)	대상 발화의 사건을 가능하게 하는 상태 혹은 사건
내적 동기(motivation)	대상 발화를 일으키는 '화자'의 감정이나 기본 욕구
감정 반응(emotional reaction)	대상 발화 사건에 대해 '청자'가 보일 수 있는 감정 반응

3. 자료 형식

데이터 세트는 제이슨(JavaScript Object Notation) 형식으로 제공되며, 각 제이슨(JSON)은 대화문, 대상 발화, 추론문의 유형, 추론문을 포함한다.

표 2는 데이터 형식의 예시이다. 주어진 훈련 데이터와 평가용 데이터는 동일한 제이슨(JSON) 형식으로 제공되며, 시험 데이터의 경우에는 각 텍스트에 대한 출력(output) 항목이 없다. 참가 팀은 해당 목록에 대해 모델의 출력 결과를 추가하여 제출한다.

※ 훈련 데이터와 제출용 데이터의 형태 및 구성 요소는 동일하다.

항목	내용
훈련용 데이터 형식의 예	<pre>[{ "id": "nikluge-2024-대화기반추론정확성평가-train-000001", "input": { "conversation": [{ "speaker": 2, "utterance": "미스터트롯 콘서트 또 취소됐네", "utterance_id": "MDRW2100011357.1.1"}, { "speaker": 1, "utterance": "코로나 때문에 콘서트 다 취소돼서 그래", "utterance_id": "MDRW2100011357.1.2"}, { "speaker": 1, "utterance": "어쩔 수 없지 뭐", "utterance_id": "MDRW2100011357.1.3"}, { "speaker": 2, "utterance": "그래도 아직 두 군데 남아 있던데 ㅋㅋ", "utterance_id": "MDRW2100011357.1.4"}, { "speaker": 1, "utterance": "그런데 코로나 때문에 취소될 것 같아", "utterance_id": "MDRW2100011357.1.5"}, { "speaker": 2, "utterance": "취소되면 아쉽겠네 ㅠㅠ", "utterance_id": "MDRW2100011357.1.6"}, { "speaker": 1, "utterance": "서울은 한 주차라도 연기해서 코로나 끝나고 했으면 좋겠어", "utterance_id": "MDRW2100011357.1.7"}, { "speaker": 2, "utterance": "콘서트 운영자들도 계획이 있겠징~", "utterance_id": "MDRW2100011357.1.8"}, { "speaker": 1, "utterance": "코로나 앞에는 답이 없당 다른 콘서트들은 하던데", "utterance_id": "MDRW2100011357.1.9"}, { "speaker": 2, "utterance": "나훈아 콘서트도 밀렸음 ㅋㅋ 뮤지컬만 하더라", "utterance_id": "MDRW2100011357.1.10"}] } }]</pre>



	<pre> {"speaker": 1, "utterance": "나훈아나 대형 가수급들은 못 하나 봐", "utterance_id": "MDRW2100011357.1.11"}, {"speaker": 1, "utterance": "소규모 공연만 가능한 듯", "utterance_id": "MDRW2100011357.1.12"}, {"speaker": 2, "utterance": "글쎄~ 빨리 코로나 끝나서 콘서트 했으면 좋겠네 ㅋ", "utterance_id": "MDRW2100011357.1.13"}, {"speaker": 1, "utterance": "미스터트롯 콘서트는 어른들 효도 콘서트나 다름없는데 할머니들 목 빠지시겠어", "utterance_id": "MDRW2100011357.1.14"}, {"speaker": 2, "utterance": "그러게~ 올해 어버이날 선물로 엄청 팔렸다면데 ㅋ", "utterance_id": "MDRW2100011357.1.15"}, }, "reference_id": ["MDRW2100011357.1.14"], "category": "원인", "inference_1": "코로나로 콘서트가 연이어 취소되고 있다.", "inference_2": "코로나로 콘서트가 연이어 개최되고 있다.", "inference_3": "코로나로 뮤지컬이 연이어 취소되고 있다." }, "output": "inference_1" }, ...] </pre> <p>- 아이디(id)와 입력(input), 그리고 출력(output)으로 구성</p>
평가용 데이터 형식의 예 (제출 전)	<pre> [{ "id": "nikluge-2024-대화기반추론정확성평가-train-000001", "input": { "conversation": [{"speaker": 2, "utterance": "미스터트롯 콘서트 또 취소됐네", "utterance_id": "MDRW2100011357.1.1"}, {"speaker": 1, "utterance": "코로나 때문에 콘서트 다 취소돼서 그래", "utterance_id": "MDRW2100011357.1.2"}, {"speaker": 1, "utterance": "어쩔 수 없지 뭐", "utterance_id": "MDRW2100011357.1.3"}, {"speaker": 2, "utterance": "그래도 아직 두 군데 남아 있던데 ㅋ", "utterance_id": "MDRW2100011357.1.4"}, {"speaker": 1, "utterance": "그런데 코로나 때문에 취소될 것 같아", "utterance_id": "MDRW2100011357.1.5"}, {"speaker": 2, "utterance": "취소되면 아쉽겠네 ㅠㅠ", "utterance_id": "MDRW2100011357.1.6"}, {"speaker": 1, "utterance": "서울은 한 주차라도 연기해서 코로나 끝나고 했으면 좋겠어", "utterance_id": "MDRW2100011357.1.7"}, {"speaker": 2, "utterance": "콘서트 운영자들도 계획이 있겠징~", "utterance_id": "MDRW2100011357.1.8"}, {"speaker": 1, "utterance": "코로나 앞에는 답이 없당 다른 콘서트들은 하던데", "utterance_id": "MDRW2100011357.1.9"}, {"speaker": 2, "utterance": "나훈아 콘서트도 밀렸음 ㅋ 뮤지컬만 하더랑", "utterance_id": "MDRW2100011357.1.10"}, {"speaker": 1, "utterance": "나훈아나 대형 가수급들은 못 하나 봐", "utterance_id": "MDRW2100011357.1.11"}, {"speaker": 1, "utterance": "소규모 공연만 가능한 듯", "utterance_id": "MDRW2100011357.1.12"}, {"speaker": 2, "utterance": "글쎄~ 빨리 코로나 끝나서 콘서트 했으면 좋겠네 ㅋ", "utterance_id": "MDRW2100011357.1.13"}, {"speaker": 1, "utterance": "미스터트롯 콘서트는 어른들 효도 콘서트나 다름없는데 할머니들 목 빠지시겠어", "utterance_id": "MDRW2100011357.1.14"}, {"speaker": 2, "utterance": "그러게~ 올해 어버이날 선물로 엄청 팔렸다면데 ㅋ", "utterance_id": "MDRW2100011357.1.15"},], "reference_id": ["MDRW2100011357.1.14"], "category": "원인", "inference_1": "코로나로 콘서트가 연이어 취소되고 있다.", "inference_2": "코로나로 콘서트가 연이어 개최되고 있다.", "inference_3": "코로나로 뮤지컬이 연이어 취소되고 있다." }, ...] </pre> <p>- 학습용 데이터와 동일한 형태 - "output" 키와 값을 제거한 데이터</p>
제출 데이터 형식의	<pre> [{ "id": "nikluge-2024-대화기반추론정확성평가-train-000001", </pre>



예	<pre> "input": { "conversation": [{"speaker": 2, "utterance": "미스터트롯 콘서트 또 취소됐네", "utterance_id": "MDRW2100011357.1.1"}, {"speaker": 1, "utterance": "코로나 때문에 콘서트 다 취소돼서 그래", "utterance_id": "MDRW2100011357.1.2"}, {"speaker": 1, "utterance": "어쩔 수 없지 뭐", "utterance_id": "MDRW2100011357.1.3"}, {"speaker": 2, "utterance": "그래도 아직 두 군데 남아 있던데 ㅋㅋ", "utterance_id": "MDRW2100011357.1.4"}, {"speaker": 1, "utterance": "그런데 코로나 때문에 취소될 것 같아", "utterance_id": "MDRW2100011357.1.5"}, {"speaker": 2, "utterance": "취소되면 아쉽겠네 ㅠㅠ", "utterance_id": "MDRW2100011357.1.6"}, {"speaker": 1, "utterance": "서울은 한 주차라도 연기해서 코로나 끝나고 했으면 좋겠어", "utterance_id": "MDRW2100011357.1.7"}, {"speaker": 2, "utterance": "콘서트 운영자들도 계획이 있겠징~", "utterance_id": "MDRW2100011357.1.8"}, {"speaker": 1, "utterance": "코로나 앞에는 답이 없당 다른 콘서트들은 하던데", "utterance_id": "MDRW2100011357.1.9"}, {"speaker": 2, "utterance": "나훈아 콘서트도 밀렸음 ㅋㅋ 뮤지컬만 하더랑", "utterance_id": "MDRW2100011357.1.10"}, {"speaker": 1, "utterance": "나훈아나 대형 가수급들은 못 하나 봐", "utterance_id": "MDRW2100011357.1.11"}, {"speaker": 1, "utterance": "소규모 공연만 가능한 듯", "utterance_id": "MDRW2100011357.1.12"}, {"speaker": 2, "utterance": "글쎄~ 빨리 코로나 끝나서 콘서트 했으면 좋겠네 ㅋㅋ", "utterance_id": "MDRW2100011357.1.13"}, {"speaker": 1, "utterance": "미스터트롯 콘서트는 어른들 효도 콘서트나 다름없는데 할머니들 목 빠지시겠어", "utterance_id": "MDRW2100011357.1.14"}, {"speaker": 2, "utterance": "그러게~ 올해 어버이날 선물로 엄청 팔렸다면데 ㅋㅋ", "utterance_id": "MDRW2100011357.1.15"},], "reference_id": ["MDRW2100011357.1.14"], "category": "원인", "inference_1": "코로나로 콘서트가 연이어 취소되고 있다.", "inference_2": "코로나로 콘서트가 연이어 개최되고 있다.", "inference_3": "코로나로 뮤지컬이 연이어 취소되고 있다.", }, "output": "inference_1" }, ...] </pre> <p>- 평가용 데이터에 “output”키와 값을 생성. - “output”에는 inference_1, 2, 3 중 적절하다고 판단되는 답안을 문자열로 입력</p>
---	--

표 2 데이터 형식의 예

표 3은 데이터의 규모이다. 데이터 세트는 훈련(train), 검증(validation), 시험(test) 데이터로 무작위 분할하여 제공한다.

	훈련	검증	시험
대화 수	758	151	605

표 3 데이터 규모

4. 기준 모델(한국어 언어 모델 블로섬 Bllossom 기반)

이 대회에의 기준 모델인 블로섬(Blossom)은 허깅페이스(huggingface)에 공개 되어있고, 과제 수행 시 기준 모델의 학습과 평가를 재현하기 위한 코드는 깃허브(github)에서 확인할 수 있다.

- [기준 모델] <https://huggingface.co/MLP-KTLim/llama-3-Korean-Bllossom-8B>
- [코드] https://github.com/teddysum/Korean_CCI_2024



5. 대회 참가 규정

- 인공지능의 한국어 능력 평가에 관심 있는 누구나 팀을 구성하여 참가할 수 있다.
- 팀 구성원은 국립국어원 언어정보나눔터 회원이어야 한다.
- 팀장은 팀원의 전자 우편 주소(회원 가입된 전자 우편 주소)를 입력하여 팀을 구성할 수 있으며, 모든 팀원이 참가 신청에 동의하면 팀 구성이 완료된다.
- 참가자는 접수 기간 중 여러 개의 팀에 참여할 수 있으며, 팀 구성 인원 제한은 없다. 다만, 접수 기간 종료 후 정성 평가 대상이 되는 팀이 여러 개라면 참가자는 과제별 하나의 팀을 소속 팀으로 결정하여야 한다.
- 참가 팀은 과제를 해결한 결과를 정해진 양식에 맞추어 제출한다(1일 5개까지 제출 가능).
- 라이선스에 문제가 없는 모델을 개발하여야 한다.
- 외부 API를 통해 호출하는 모델(OpenAI API 등)은 제출할 수 없다.
- 개발한 모델은 RTX 4090 24GB 1개에서 구동 가능하여야 한다.
- 외부에 공개된 데이터는 사용이 불가하며 과제 데이터(말뭉치)를 초거대언어모델(LLM) 등을 활용해 변형 및 증강하여 사용할 수 없다. 모델 입력을 위한 형식 및 형태 변형만이 허용된다.
- 순위표 기준 정성 평가 신청서를 제출한 상위 5개 팀을 정성 평가 및 발표 평가 대상으로 선정한다. (해당 팀에게는 별도 안내 예정)
- 최종 수상자는 정량 평가, 정성 평가, 발표 평가를 종합하여 선정한다.

6. 시스템 사용 방법

- 인공지능(AI)말평 → 알립니다 → 사용 안내 메뉴 내 동영상 참고

7. 문의

- 진행 중 과제 → 대화 맥락 추론 클릭 → 문의(들여가기 후 작성 가능)