

2024.03.05 회의록

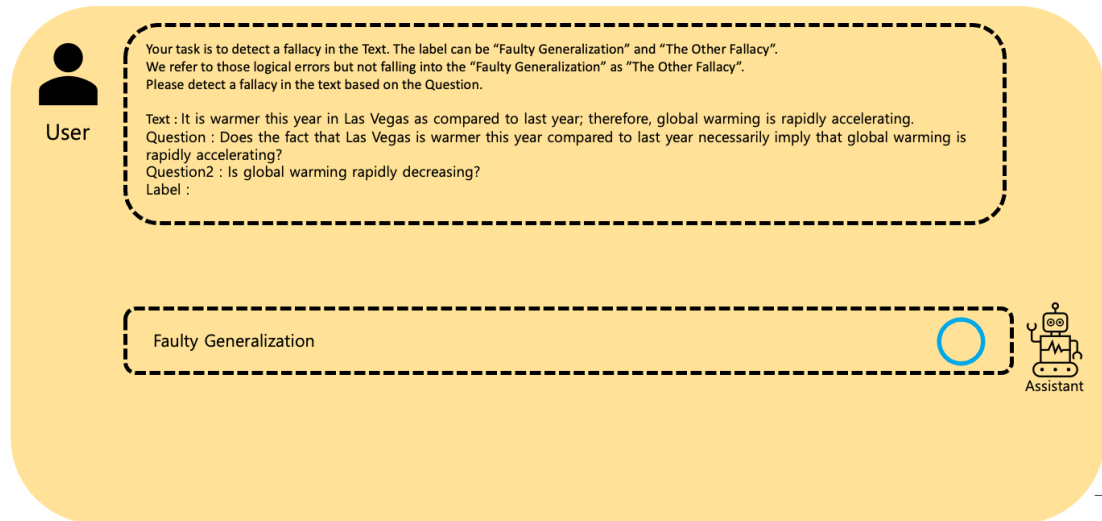
금일 회의 내용

1. Faulty Generalization 클래스에 대해서 집중적으로 분석을 진행함
 - a. 분석을 통해 일반화 오류를 가지고 있는 문장에 대해서 질문을 형성함
2. Faulty Generalization 문장들을 확인해서 $A < Q$, $A > Q$ 타입에 따른 규칙을 찾고 질문 형태를 선택함
 - a. **$A < Q$** : 단순히 Q가 사실인지를 되묻는 형태의 질문이면 된다. 왜냐하면 이 문장은 소수(한 명, 친구, 가족)에서 주장하는 내용(A)을 대중적인 관념(Q)을 기반으로 사용하기 때문에, 근본적으로 Q가 사실인지를 직접적으로 물으면 논리 오류를 해결할 수 있음.
 - b. **$A > Q$** : 경험한, 주위 사람에게 들은, 본 내용(Q)을 기반으로 일반화 적인 주장(A)를 하는 경우로, 이런 문장 같은 경우에는 A와 Q의 관계를 묻거나(Question), A를 되묻는 형태의 질문(Question2)을 하면 된다.
3. prompting을 통해서 $A > Q$ 문장들은 Question, Question2를 생성하고, $A < Q$ 문장은 직접 만듦.
 - a. Question : A와 Q의 관계를 묻는 형태로 일반화된 주장에 대한 논리적 결함을 직접적으로 다룸
 - b. Question2 : A를 묻는 경우로 주장된 결과를 의심하거나 부정할 수 있도록 유도하는 형태
4. 실험은 faulty generalization을 맞추지 못할 경우, Question(or Question2)으로 다시 물어보는 형태로 실험을 진행함
5. 추가실험 : Text로부터 만든 Question의 영향성을 파악하기 위해, 정답(레이블)을 틀릴 시, Question을 사용하지 않고, 단순히 rethink하라는 prompt만 넣었을 때의 성능을 확인함 -> Rethink가 성능 향상을 줬을 수도 있어서, Rethink만 하는 실험을 추가함.

피드백 및 나왔던 내용

1. 실험 세팅에서 잘못된 것 같다 : 틀렸을 경우에만 진행하는 것은 정답을 알려주는 것이다.

a. 재실험을 진행함.



User

Your task is to detect a fallacy in the Text. The label can be "Faulty Generalization" and "The Other Fallacy". We refer to those logical errors but not falling into the "Faulty Generalization" as "The Other Fallacy". Please detect a fallacy in the text based on the Question.

Text : It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.

Question : Does the fact that Las Vegas is warmer this year compared to last year necessarily imply that global warming is rapidly accelerating?

Question2 : Is global warming rapidly decreasing?

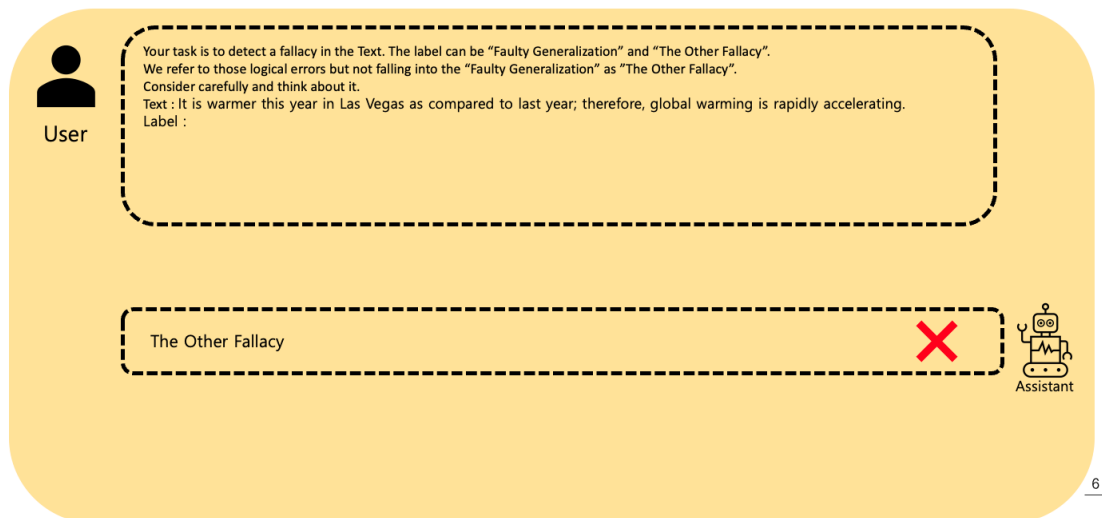
Label :

Faulty Generalization

Assistant

5

b. Rethink 실험이 있었는데 이것은 단순히, 기존 문장 + "심사숙고해서 생각해봐" 의 형태로 변경해서 실험함 → 이 실험은 질문형태가 아닌 다시 생각해봐라 했을 때의 성능이 올라가는지 확인하기 위해서 진행함



User

Your task is to detect a fallacy in the Text. The label can be "Faulty Generalization" and "The Other Fallacy". We refer to those logical errors but not falling into the "Faulty Generalization" as "The Other Fallacy". Consider carefully and think about it.

Text : It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.

Label :

The Other Fallacy

Assistant

6

2. 우선, Faulty generalization에 대해서만 진행했는데, 이것만으로는 논문을 쓰기에는 약하다. 다른 클래스에 대해서도 질문 형태를 만드는 것을 해봐야할 것 같다.

- a. False causality, Irrelevant Authority, Post Hoc, Cherry Picking과 같은 지식 그래프가 사용되면 좋을 것 같다고 생각했던 클래스에 대해서도 질문을 만들다보면 **General한 질문**을 만드는 방식을 찾을 수 있지 않을까?
- b. 지식 그래프가 사용되면 좋을 것 같은 클래스들의 공통점은 문장 내 핵심 키워드, 주장 간의 연결관계의 부족 및 잘못된 연결관계였기 때문에 이 관계를 묻는 질문을 만든다면??
- c. a,b 내용을 확인하기 위해 **False Causality**를 Faulty Generalization 처럼 분석하고 찾아보라고 하심. 분석하다보면 General하게 적용될 수 있는 규칙이 있을 것이며, 추후에 further analysis에 클래스별 더 적절한 질문 형태로 했을 때의 성능도 보여주는 방식으로 하면 어떨까?

재실험 결과

Type/Metric	Total Accuracy	Precision	Recall	F1
No Question	0.9(±0.01)	0.33(±0.03)	0.3(±0.01)	0.32(±0.02)
Question	0.92(±0.02)	0.5(±0.07)	0.78(±0.04)	0.61(±0.06)
Question2	0.92(±0.01)	0.47(±0.04)	0.76(±0.02)	0.58(±0.02)
Carefully think	0.9(±0.002)	0.35(±0.02)	0.32(±0.04)	0.33(±0.03)