

Scenarios and Approaches for Situated Natural Language Explanations

Evaluation

- Sentence similarity
 - Calculate similarity between ground truth explanation (h_c) and LLM generated (e_j) explanation
- Cross-entropy loss -> matching score
 - Convert similarities to probabilities, then calculate loss

$$p_{cj} = \frac{\exp(\text{sim}(h_c, e_j))}{\sum_{c=1}^3 \exp(\text{sim}(h_c, e_j))}$$

$$\text{Matching}_j = - \sum_{c=1}^3 y_c \log(p_{cj})$$

Scenarios and Approaches for Situating Natural Language Explanations

Result

- LLMs can generate prompts that result in explanations more precisely aligned with the target situations
- "You are a helpful assistant..." is not a necessary prompt technique for situated NLE tasks
- In-context learning prompts only can help LLMs learn the demonstration template but can't improve their inference performance