

Plan and Solve Prompting: Improving Zero-Shot COT Reasoning by LLM

Sparsh Badjate,
Soham Dahane,
Manav Patel,
Dhruv Gattani,

29/11/2024

Problem Statement

- **Input:** Math word problem + Prompting Techniques
- **Output:** Final Answer to the given word problem
- **Example:**
 - **Input:** Q:Charlie has three times as many Facebook friends as Dorothy. James has four times as many friends on Facebook as Dorothy. If Charlie has 12 friends on Facebook, how many Facebook friends does James have?

A:Let's first understand the problem, extract relevant variables and corresponding numerals, and devise a complete plan. Then, let's carry out the plan, calculate intermediate variables (pay attention to correct numerical calculation and commonsense), solve the problem step by step, and show the answer.
 - **Output:** Answer is: 16.0 + Chain of thought

Motivation

- **Scaling Alone Isn't Enough:** Increasing model size doesn't guarantee better performance on complex reasoning tasks.
- **Power of Prompts:** Effective prompts are key to unlocking the full potential of LLMs, guiding them to perform better on tasks.
- **Improving Reasoning:** Well-designed prompts can enhance multi-step reasoning and task accuracy in LLMs.
- **Efficiency Through Prompts:** Prompt engineering maximizes model performance without the need for extensive retraining or scaling.

Literature Review

Large Language Models are zero shot reasoners

This paper introduces Zero-shot Chain of Thought (CoT) prompting, a method that significantly improves large language models' (LLMs) reasoning abilities by adding a simple phrase like "Let's think step by step" to prompts. Unlike few-shot CoT, which requires task-specific examples, The study highlights the potential of task-agnostic prompting to unlock broad cognitive abilities in LLMs.

Plan and Solve Prompting: Improving Zero-Shot CoT Reasoning by LLM

The paper advances zero-shot reasoning by introducing Plan-and-Solve Prompting (PS+), which addresses limitations in CoT and PoT, such as error handling and reasoning quality. By incorporating planning and detailed instructions in prompts, it enhances the accuracy and robustness of zero-shot reasoning, reducing the need for manual demonstrations.

Dataset

- Dataset instance- GSM8K: High-quality grade-school math word problems requiring multi-step reasoning
- # Samples: 1,319
- Avg. Words per Instance: 46.9
- Answer Type: Numerical

Existing Prompting Methods

- **Zero-Shot CoT** : Guides the model to perform step-by-step reasoning using natural language cues (for eg. appending “Let’s think step by step”) without providing any prior examples.
- **Few-shot CoT** : Demonstrates step-by-step reasoning by providing a few examples before asking the model to solve a new problem.

Prompting Methods

- **Zero-shot CoT with self consistency** : Combines zero-shot step-by-step reasoning with multiple sampled reasoning paths to select the most consistent answer based on majority voting.
- **Zero-shot complexity based self consistency** : This is our method where we select the most consistent answer by sampling multiple reasoning paths, with preference given to paths demonstrating higher complexity or depth in reasoning.

Results

Prompting Technique	Correct	Incorrect
Let's Think Step by Step	1286	33
Plan and Solve	1303	16
Plan and solve with SC	1307	12
Modified SC	1311 - 1307	8(best case) - 12(worst case)

Analysis

Question: A food caterer was told to prepare gourmet hot dogs for 36 guests. While most people would only eat one hotdog, he prepared enough for half of the guests to be able to have two hotdogs. However, 40 guests showed up, and everyone wanted a second hotdog. How many guests did not get a second hotdog?

Answer : 26

Raw Prompting

The caterer prepared enough hot dogs for half of the 36 guests to have two hot dogs each. Calculating the number of hot dogs prepared:- Half of 36 guests = $36 / 2 = 18$ guests- Each of these 18 guests is supposed to have 2 hot dogs, **so the total number of hot dogs prepared = $18 * 2 = 36$** hot dogs. Now, 40 guests showed up, and everyone wanted a second hot dog. Calculating the total number of hot dogs needed:\n- If each of the 40 guests wants 2 hot dogs, the total number of hot dogs needed = $40 * 2 = 80$ hot dogs. Now, we compare the number of hot dogs prepared with the number needed:\n- Hot dogs prepared = 36\n- Hot dogs needed = 80 .To find out how many guests did not get a second hot dog, we first determine how many guests can be served with the 36 hot dogs- Each guest can have 1 hot dog, so 36 guests can be served 1 hot dog each.\n- This leaves $40 - 36 = 4$ guests who did not get a hot dog at all. Now, we need to see how many guests can get a second hot dog with the remaining hot dogs:\n- After serving 36 guests 1 hot dog each, there are 0 hot dogs left for seconds.\n\nThus, since there are no hot dogs left for seconds, all 40 guests who wanted a second hot dog did not get one. Therefore, the number of guests who did not get a second hot dog is ****40****.

Solve Step by step

"Let's break down the problem step by step:

1. **Initial Guest Count:** The caterer was told to prepare for 36 guests.
2. **Hot Dog Preparation:** The caterer prepared enough hot dogs for half of the guests to have two hot dogs each. - Half of 36 guests is $\left(\frac{36}{2} = 18\right)$ guests. - If these 18 guests each have 2 hot dogs, **the total number of hot dogs prepared is $(18 \times 2 = 36)$ hot dogs.**
3. **Actual Guest Count:** However, 40 guests showed up.
4. **Hot Dog Demand:** Since everyone wanted a second hot dog, we need to calculate how many hot dogs are needed for all 40 guests:
- If each of the 40 guests wants 2 hot dogs, the total number of hot dogs needed is $(40 \times 2 = 80)$ hot dogs.
5. **Hot Dogs Available:** The caterer only prepared 36 hot dogs.
6. **Hot Dogs Shortage:** To find out how many guests did not get a second hot dog, we need to determine how many guests can be served with the 36 hot dogs:
- Each guest can have 1 hot dog first, which accounts for 40 hot dogs. After that, we can see how many can get a second hot dog with the remaining hot dogs.
- After serving 1 hot dog to each of the 40 guests, there are 36 hot dogs used, leaving 0 hot dogs for a second serving.
7. **Conclusion:** Since there are no hot dogs left after the first serving, none of the guests can receive a second hot dog. Therefore, all 40 guests wanted a second hot dog but did not get one.
Thus, the number of guests who did not get a second hot dog is ****40**.**

Plan and Solve

START

Step 1: Initial Preparation

total_guests_expected = 36

hot_dogs_prepared = (total_guests_expected / 2) * 2 # Prepare enough for half of the guests to have two hot dogs

Step 2: Guests Who Showed Up

total_guests_arrived = 40 # 40 guests showed up

Step 3: Total Hot Dogs Needed

hot_dogs_needed = total_guests_arrived * 2 # Each guest wants 2 hot dogs

Step 4: Comparing Prepared Hot Dogs to Needed Hot Dogs

hot_dogs_short = hot_dogs_needed - hot_dogs_prepared # Calculate shortage

Step 5: Guests Missing Second Hot Dog

guests_missing_second_hot_dog = total_guests_arrived - hot_dogs_prepared # Guests missing a second hot dog

Step 6: Final Answer

PRINT "Number of guests who did not get a second hot dog:", guests_missing_second_hot_dog

END

Plan and Solve with SC

START

Step 1: Initial Plan for Hot Dogs

total_guests_expected = 36 # Guests initially planned

half_guests = total_guests_expected / 2 # Half of the guests expected to have two hot dogs

hot_dogs_prepared = half_guests * 2 # Total hot dogs prepared

Step 2: Actual Guests and Needs

total_guests_arrived = 40 # Actual guests who showed up

hot_dogs_needed = total_guests_arrived * 2 # Total hot dogs needed (2 per guest)

Step 3: Compare Hot Dogs Prepared and Needed

hot_dogs_short = hot_dogs_needed - hot_dogs_prepared # Hot dog shortage

Step 4: Guests Missing a Second Hot Dog

guests_missing_second_hot_dog = total_guests_arrived - hot_dogs_prepared # Guests missing a second hot dog

Step 5: Final Answer

PRINT "Number of guests who did not get a second hot dog:", guests_missing_second_hot_dog

END

Plan and Solve with complexity SC

"To tackle this problem, let's break it down step by step using relevant variables.\n\n1. ****Understanding guest requirements****:\n - Total guests expected = 36.\n - Most guests would eat 1 hotdog, so if all had just one, that would be 36 hotdogs.\n - However, the caterer prepared enough for half of the guests to have a second hotdog.\n - Half of 36 = 18 guests can have 2 hotdogs each.\n - **Total hotdogs prepared = 36 (one for each guest) + 18 (for the second hotdogs) = 54 hotdogs prepared.**\n\n2. ****Situation when more guests show up****:\n - Actual guests who showed up = 40.\n - All 40 guests want a second hotdog.\n\n3. ****Calculating Hotdog Demand****:\n - Each of the 40 guests wants 2 hotdogs, so total hotdogs needed = $40 * 2 = 80$ hotdogs.\n\n4. ****Comparing Prepared Hotdogs to Demanded Hotdogs****:\n - Hotdogs prepared = 54.\n - Hotdogs demanded = 80.\n - Thus, the shortfall = $80 - 54 = 26$ hotdogs.\n\n5. ****Determining how many guests did not get a second hotdog****:\n - The caterer served each of the 40 guests their first hotdog.\n - After this, we have 54 hotdogs served (for all 40 guests having their first).\n - Remaining hotdogs = $54 - 40 = 14$ hotdogs.\n - The guests who can get a second hotdog = 14 (since everyone wants a second hotdog).\n - Therefore, the guests who cannot get a second hotdog = Total guests - Guests who get a second hotdog = $40 - 14 = 26$ guests.\n\n****Final Answer****: 26 guests did not get a second hotdog."

Question

Michael is replacing the carpet in his bedroom. The new carpet he's chosen costs \$12 per square foot, plus \$2 per square foot for padding underneath. His contractor charges \$4 per square foot to remove the old carpet, and \$34 per square foot to install the new carpet. His bedroom measures 18 feet by 12 feet. How much will it cost Michael to replace the carpet?

Answer: 11232

Raw Prompt

START

Step 1: Calculate the area of the bedroom

length = 18 # in feet

width = 12 # in feet

area = length * width # Area = Length × Width

Step 2: Calculate the cost of the new carpet

cost_per_sqft_carpet = 12 # in dollars/square foot

cost_carpet = area * cost_per_sqft_carpet # Cost of carpet = Area × Cost per square foot

Step 3: Calculate the cost of padding

cost_per_sqft_padding = 2 # in dollars/square foot

cost_padding = area * cost_per_sqft_padding # Cost of padding = Area × Cost per square foot

Step 4: Calculate the cost to remove the old carpet

cost_per_sqft_removal = 4 # in dollars/square foot

cost_removal = area * cost_per_sqft_removal # Cost to remove old carpet = Area × Cost per square foot

Step 5: Calculate the cost to install the new carpet

cost_per_sqft_install = 34 # in dollars/square foot

cost_install = area * cost_per_sqft_install # Cost to install = Area × Cost per square foot

Step 6: Calculate the total cost

total_cost = cost_carpet + cost_padding + cost_removal + cost_install # Total cost

Output the final result

PRINT "Total cost for Michael to replace the carpet is:", total_cost

END

Plan and Solve

START

Step 1: Calculate the area of the bedroom

```
length = 18 # feet
width = 12 # feet
area = length * width # Area = 18 * 12 = 216 square feet
```

Step 2: Calculate the cost of the new carpet

```
cost_per_sqft_carpet = 12 # dollars/square foot
cost_carpet = area * cost_per_sqft_carpet # Cost of carpet = 216 * 12 = 2592 dollars
```

Step 3: Calculate the cost of padding

```
cost_per_sqft_padding = 2 # dollars/square foot
cost_padding = area * cost_per_sqft_padding # Cost of padding = 216 * 2 = 432 dollars
```

Step 4: Calculate the cost to remove the old carpet

```
cost_per_sqft_removal = 4 # dollars/square foot
cost_removal = area * cost_per_sqft_removal # Cost of removal = 216 * 4 = 864 dollars
```

Step 5: Calculate the cost to install the new carpet

```
cost_per_sqft_installation = 34 # dollars/square foot
cost_installation = area * cost_per_sqft_installation # Cost of installation = 216 * 34 = 7344 dollars
```

Step 6: Calculate the total cost

```
total_cost = cost_carpet + cost_padding + cost_removal + cost_installation # Total cost = 2592 + 432 + 864 + 7344 = 11332 dollars
```

Output the result

```
PRINT "Total cost for replacing the carpet is: $", total_cost # Output: $11,332
```

END

Plan and solve with SC

"To determine the total cost for replacing the carpet in Michael's bedroom, we need to break down the costs into several components based on the details provided.\n\n

Step 1: Calculate the area of the bedroom\nThe area (A) of the bedroom can be calculated using the formula for the area of a rectangle:\n\n $A = \text{length} \times \text{width}$ \n\nGiven:\n- Length = 18 feet\n- Width = 12 feet\n\nNow, calculate the area:\n $A = 18 \text{ ft} \times 12 \text{ ft} = 216 \text{ sq ft}$ \n\n

...

Step 4: Calculate the cost to install the new carpet\nThe contractor charges \$34 per square foot for the installation of the new carpet:\n\n $\text{Installation cost} = 216 \text{ sq ft} \times 34 \text{ dollars/sq ft} = 7344 \text{ dollars}$ \n\n### Step 5: Calculate the total cost\nNow, we can sum up all the costs to find the total cost for replacing the carpet:\n\n $\text{Total cost} = \text{Removal cost} + \text{Total carpet and padding cost} + \text{Installation cost}$ \n\nSubstituting the values we calculated:\n\n $\text{Total cost} = 864 \text{ dollars} + 3024 \text{ dollars} + 7344 \text{ dollars} = 11232 \text{ dollars}$ \n\n### Final Answer\nThe total cost for Michael to replace the carpet in his bedroom will be **\$11,232**.",

Improvements over the paper

- We implemented a complexity-based self-consistency prompting approach to prioritize reasoning steps.
- It is a modified ensemble technique which uses the predictions with maximum number of COT to perform ensemble on to predict output
- This method demonstrated improved performance, in solving difficult math word problems requiring multi-step reasoning.
- The approach is intuitive, leveraging the fact that more COT means better problem solving

Learnings

- Analyzing how different prompting strategies influence the effectiveness and accuracy of solutions across various problem-solving tasks.
- Exploring the role of API keys in enhancing functionality and access to external services for seamless integration during coding tasks.
- Highlighting the crucial role that structured reasoning plays in improving the reliability and depth of responses in CoT-based problem-solving approaches.

Evaluation Scheme

- Demo working- ((10) if not working or no GUI - (0))
 - Literature understanding- (10)
 - Analysis- (10)
 - Improvement over paper- (10)
-
- **Note: Must have GUI, otherwise no mark will be given for demo.**