PromptGen

Divyansh Rajesh Jain, Varsha Sivaprakash

PromptGen: Problem Statement

Research Question:

Given a set of domain specific constraints, can we generate a large dataset of natural language question and answer pairs that can be used to verify or finetine an LLM?

Relevant Prior Work

VeriPlan:

A constraint based system for LLM auto self-verification for scheduling tasks. The user writes constraints, based on which the model schedules tasks and checks whether they meet constraints. Users on the loop to ensure correctness.

BERT Score:

Evaluation metric to compare natural language sentences semantically, ranging from 0-1, where higher the score means the more semantically close.

Preliminary Experiment Approach

- Defined a set of constraints in a Prolog based system to declaratively generate question/answer datasets based on those constraints. The constraints were about United States geography
- Evaluated over 400 prompts on the LLM, generated over 30000 prompts
- Evaluation metrics:
  - BLEU scores and BERT precision, recall and F1 against baseline GeoBase
  - LLM evaluation on Llama 2 7B model

Preliminary Results 1: BLEU and BERT

```
"category": "Border Question",
"reference_question": "what states border montana?",
"system_generated_question": "What states does montana border?"


"category": "Capital Question",
"reference_question": "what is the capital of washington?",
"system_generated_question": "What is washington's capital?"


"category": "Population Question",
"reference_question": "how many citizens in alabama?",
"system_generated_question": "What is alabama's population?"


"category": "River in State Question",
"reference_question": "what rivers are in utah?",
"system_generated_question": "What rivers does utah contain?"


"category": "Major Cities in State Question",
"reference_question": "what are the major cities of texas?",
"system_generated_question": "What major cities does texas contain?"
```

Category	BLEU Score	BERT Precision	BERT Recall	BERT F1
Border Question	11.48	0.9394	0.9573	0.9482
Capital Question	9.42	0.9690	0.9682	0.9686
Population Question	8.75	0.9263	0.9268	0.9265
River in State Question	10.68	0.9529	0.9529	0.9529
Major Cities in State Question	13.54	0.9344	0.9433	0.9388

Preliminary Results 2: LLM Evaluation

Category	# Q's Generated	# Q's Asked	# Q's Correct	% Passed
Geography US State Abbreviation Knowledge	51	51	50	98.04%
Geography US State Capital Knowledge	51	51	49	96.08%
Geography US Rivers Knowledge	46	46	4	8.70%
Geography US State Simple Border Questions	49	49	3	6.12%
Geography US States Complex Border Questions AND	534	100	47	47.00%
Geography US States Complex Border Questions AND BUT	23601	100	0	0.00%
Geography US States Complex Total population	19600	100	0	0.00%

LLM Evaluation Sample System Prompt

Category: Geography US Rivers Knowledge

System prompt: Please respond with answers formatted exactly like this (without quotes): the [River Name] river flows through [List of states river flows through in alphabetical order with an Oxford comma]. Replace [State] with the state mentioned in the prompt, and [List of states river flows through] with the actual states that the river flows through. Do not include any brackets, quotation marks, or other symbols in your answer. Respond with a single sentence only.

What do these results mean?

Example 1: LLM made a genuine mistake

```
Question: What is california's abbreviation?
LLM      : ohio's abbreviation is oh.
Correct : california's abbreviation is ca.
```

```
Question: what is Georgia, Iowa, and Pennsylvania's population in total?
LLM      : georgia, iowa, and pennsylvania's population in total is 10.6 million.
Correct : georgia, iowa, and pennsylvania's population in total is 27701660.
```

```
Q: which state do Nebraska, South Dakota, and Utah neighbor in common?
LLM A: nebraska, south dakota, and utah neighbor the state of montana in common.
Correct A: nebraska, south dakota, and utah neighbor wyoming in common.
```

Example 2: Partially Correct Marked Wrong

```
Question: What states does the st. francis river flow through?
LLM      : The St. Francis River flows through Arkansas, Illinois, Missouri, and Tennessee.
Correct  : the st. francis river flows through arkansas and missouri.
```

```
Question: What is district of columbia's capital?
LLM      : washington d.c.'s capital is washington d.c.
Correct  : district of columbia's capital is washington.
```

Example 3: Marked wrong even though correct

```
Question: which state do Nevada, Oregon, and Wyoming neighbor in common?
LLM      : nevada, oregon, and wyoming all neighbor the state of idaho in common.
Correct  : nevada, oregon, and wyoming neighbor idaho in common.
```

```
Question: What states does the wateree catawba river flow through?
LLM      : The Wateree-Catawba River flows through South Carolina and North Carolina.
Correct  : the wateree catawba river flows through north carolina and south carolina.
```

```
Question: What is minnesota's capital?
LLM      : minnesota's capital is saint paul.
Correct  : minnesota's capital is st. paul.
```

Thank You!