CS-Tab-CoT: Zero-Shot Tabular Chain of Thought for Common Sense Reasoning Tasks

Kaya Gouin
School of Computer Science
Carleton University
Canada

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

Pretrained Large Language Models (LLMs) are great at NPL tasks. They got better in recent years in terms of accuracy. Pretrained Large Language Models are models which estimate probability distributions over text. In recent years, they have shown remarkable success on a wide range of natural language processing tasks. Large language models are regarded as excellent few-shot learners when given task-specific exemplars, with their success stemming from these few-shot learning abilities. Large language models have been argued to be decent zero-shot reasoners when provided with a simple phrase that encourages step-by-step reasoning [kojima2022]. Conditioning LLMs is referred to as prompting. System-1 vs system-2 tasks.

A common theme with existing methods, as we'll see, is that common sense reasoning is especially hard for LLMs. Techniques that have been put forward to better condition LLMs are often focused on arithmetic tasks and other tasks that require careful systematic reasoning. Common sense reasoning is different. It is an inherently human trait. Requires experience, intelligence, and creativity. The way in which humans reason about common sense reasoning tasks is highly variable. This depends on the specific task, but across the board humans don't really reason in a step-by-step fashion when it comes to common sense reasoning. It's more of a feeling, or an obvious answer based on previous experience and general knowledge of the world. Again, intelligence, creativity, and even imagination are key aspects of our common sense reasoning skills.

In this work, we follow the work of Jin and Lu and we suggest a novel way of conditioning LLMs... We present an augmented tabular-format chain-of-thought prompting method designed to elicit high performance on common sense reasoning tasks. And beyond seeking high performance, we're interested in knowing how different types of reasoning instructions (don't like how this is said) elicit different reasoning paths. We test it against established baselines. We compare it to a greater range of common sense reasoning tasks as compared to the previous works we'll be talking about (if we have the chance to do so).

We begin with some background on recent improvements in methods of conditioning Large Language Models, ...

2 Background

Here we'll go over key methods of prompting LLMs. Talk about standard few- and zero-shot prompting methods here or somewhere before.

2.1 Few-Shot-CoT

By Wei et al. from 2022. They mainly look at few-shot CoT prompting. Explain the origin of this 'chain of thought' prompting method. Explain their model. This is where the chain-of-thought prompting method was first presented (I think. Double check). Following their work, Wang et al. proposed a slightly modified version, whereby ...

2.2 Zero-Shot-CoT

By Kojima et al. One or two paragraphs for each subsection within this section explaining the general idea behind the method. Explain their model, and the improvements over few-shot-cot. Here the authors wanted a more generalizable method, a task-agnostic method, and also wanted to show the untapped potential of Zero-Shot-CoT.

2.3 Zero-Shot-Least-to-Most

By Zhou et al. Explain how this one works, similar to what we did in previous subsections. Explain their model, and the improvements over previous work. Here they say that their method of conditioning is not effective for common sense reasoning tasks. That being said, it's the general idea of decomposition and subproblem solving that was the key innovation. Here they wanted a generalizable approach whereby a language model would be better able to solve easy-to-hard generalizations. They look at few-shot and zero-shot contexts.

2.4 Zero-Shot-Tab-CoT

By Jin, Lu. Explain how this one works, similar to what we did in previous subsections. Explain their model, and the improvements over previous work. Here they were looking at a highly organized way of structuring the reasoning process in a zero-shot context.

3 CS-Tab-CoT

```
|step|subquestion|process|result|
```

```
|deductive reasoning process|result|
|inductive reasoning process|result|
|abductive reasoning process|result|
|logical reasoning process|result|
|critical reasoning process|result|
|counter-factual reasoning process|result|
|analogical reasoning process|result|
|inferential reasoning process|result|
|cause-for-effect reasoning process|result|
|creative reasoning process|result|
```

No 'step' and no 'subquestion' part, since common sense reasoning tasks aren't structured in such a way to promote easy decomposition into simpler subquestions.

3.1 Tasks and Datasets

Here we look at the two datasets used in some previous works presented: CommonSenseQA [Talmor et al., 2019] and StrategyQA [Geva et al., 2021]. I we have the chance, we might look at some extra ones from `https://commonsense.run/datasets/`.

3.2 Large Language Models

text-davinci-002 and code-davinci-002 are the main ones we'll be using so that we can easily compare our results to previous work.

3.3 Baselines

Here we're mainly interested in comparing methods for zero-shot contexts. We'll look at standard, CoT, and Tab-CoT (all of these are in zero-shot contexts). Insert a few tables here.

3.4 Answer Extraction

Same answer extraction technique from Jin and Lu.

4 Results and Discussion

Compare our work to zero-shot, zero-shot-cot, and tab-cot.

5 Conclusion and Future Work

References

[1] Ziqi Jin and Wei Lu. Tab-CoT: Zero-shot Tabular Chain of Thought, 2023. arXiv:2305.17812.

[2] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, Yusuke Iwasawa. Large language models are zero-shot reasoners, 2022. arXiv:2205.11916.

[3] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models, 2022. arXiv:2203.11171.

[4] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. Chain of thought prompting elicits reasoning in large language models, 2022. arXiv:2201.11903.

[5] Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc Le, Ed Chi. Least-to-most prompting enables complex reasoning in large language models, 2022. arXiv:2205.10625.