Review of: ""Let's Argue Both Sides": Argument Generation Can Force Small Models to Utilize Previously Inaccessible Reasoning Capabilities"

Tianbo Ji[1]

1 Nantong University, Nantong, China

Potential competing interests: No potential competing interests to declare.

This paper introduces a novel prompting technique, aimed at improving model performance in logical reasoning tasks. Through empirical evaluation, the authors demonstrate the superiority of this method across multiple datasets and highlight its significant benefits in smaller models. However, I believe that the article needs further refinement in certain areas to ensure comprehensive and accurate content.

Comments

1. **Evaluation of Method Effectiveness**

   The article mentions that the effectiveness of Argument Generation does not seem to be good and even has negative optimization cases. The authors are advised to provide more specific data and analysis of these negative effects to ensure readers can fully understand the actual performance of the method.
2. **Definition of Performance Differences**

   The specific meaning of performance differences mentioned in the article should be clarified with more explanation. The authors are advised to clearly define the performance differences and provide specific metrics and data support.
3. **Reasonableness of Method Comparison**

   The performance of Argument Generation is always compared with the worst method, which seems unreasonable. The authors are advised to compare more with CoT (Chain of Thought) and explain the differences and advantages and disadvantages.
4. **Limitations of the Method**

   The article mentions that the dependence of Argument Generation on the predefined number of options for parameters is an inherent limitation. The authors are advised to further explore how to overcome this limitation and provide possible solutions.
5. **Coverage of Datasets**

   Although the authors have made every effort to cover datasets related to evaluating various model functions, other task-independent prompting methods may outperform this method in many untested metrics. The authors are advised to further expand the coverage of datasets to ensure the universality of the method.
6. **Relationship Between Model Size and Prompting Methods**

   The article mentions that the method yields greater benefits in smaller models. The authors are advised to further

explore the relationship between model size and prompting methods and provide more theoretical support and empirical data.

In conclusion, the research is sound and well-motivated, and the paper is easy to follow. I suggest a minor revision, and the paper can benefit from revising according to the aforementioned comments.