ItD: Large Language Models Can Teach Themselves Induction through Deduction

Wangtao Sun, et al

Institute of Automation, Chinese Academy of Sciences

March 9, 2024

Abstract

- LLMs have limited ability for induction
  - Recent researches suggested "Post process"
  - But their performance is still limited

- New framework: Induction through Deduction
  - Deductive Data Generation module
  - Naive Bayesian Induction module

Hypothesis search & Refinement

- One of recent works
  - Weak inductor performs several inductions
  - Deductive verification and refinement

- Inherent induction ability is still limited
 - The inductor is weak.
  - LLMs are usually good at deduction.

(a) Hypothesis Search & Refinement
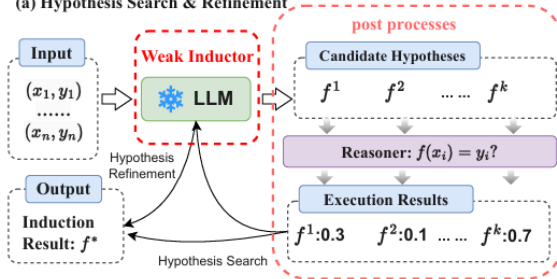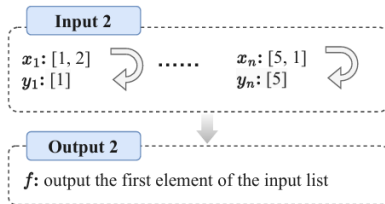
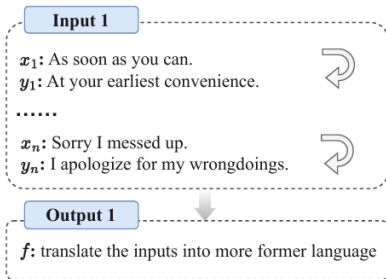# Induction through Deduction

- Deductive Data Generation module
  - Using given samples, generates more samples
  - Combined with LoRA, inductor performs induction

- Naive Bayesian Induction module
  - Verify reliability of the generator's augmentation
  - Fine-tunes LLM to predict $p(f|x, y)$ supposing (x_i, y_i) are iid.

(b) ItD
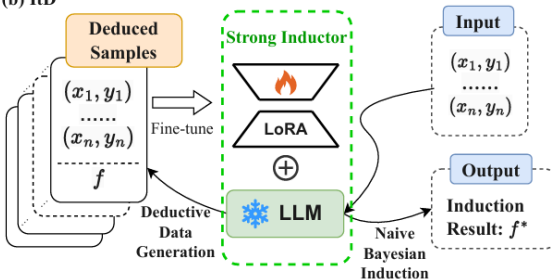
Experiment

- Instruction Induction Task (Semantic)
  - x and y are Natural Language sentences

- List Function Task (Symbolic)
  - x and y are list with numeric data

Input 1

x_1: As soon as you can.
y_1: At your earliest convenience.

......

x_n: Sorry I messed up.
y_n: I apologize for my wrongdoings.

Output 1

f: translate the inputs into more former language

Input 2

x_1: [1, 2] x_n: [5, 1]
y_1: [1]   ...... y_n: [5]

Output 2

f: output the first element of the input list

Conclusion

- ItD is superior to existing methods
  - ItD achieved improvement of 36%, 10% with previous SOTA
  - Improvement of both the semantic, symbolic deduction

- Limitation
  - Compared to semantic task, symbolic performance was not satisfying
  - Naive Bayesian is greedy, leading to local-optima