Making Large Language Models Better Reasoners with Step-Aware Verifier

Yifei Li, Zeqi Lin, Shizhuo Zhang, Qiang Fu, Bei Chen

Jian-Guang Lou, Weizhu Chen

Motivation

Making Large Language Models Better Few-Shot Reasoners

with Diverse Verified Reasoning Steps

"All Roads Lead to Rome"

"The truth is not necessarily in the hands of the majority"

"Reasoning is a multistep process"

1. *Wisdom of the crowd*

- We need induce more diverse reasoning paths from the language model

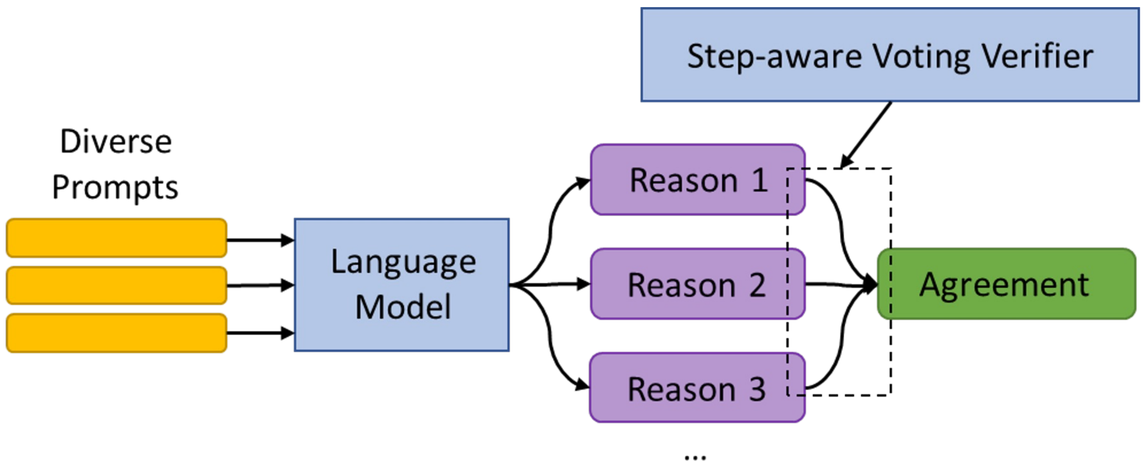
2. *Reflective thinking*

- Not all reasoning paths are equally good
- We need distinguish good reasoning paths from bad reasoning paths

3. *Multistep thinking*

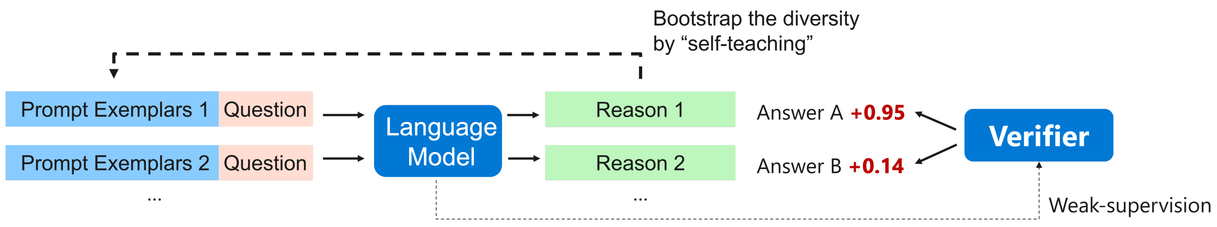
- Each reasoning path consists of multiple steps
- We need to look into the steps, rather than deal with all steps of a reasoning path in a whole

Our Method: DIVERSE



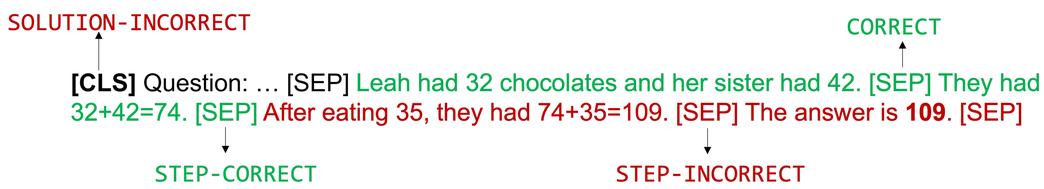
- **Diverse Reasoning Paths:** Diverse prompts + Temperature Decoding
- **Voting Verification:** Reasoning paths weighted-voting with verifier scores
- **Step Correctness:** Obtain step-level labels to achieve a step-aware verifier

Diverse Prompts & Voting Verification

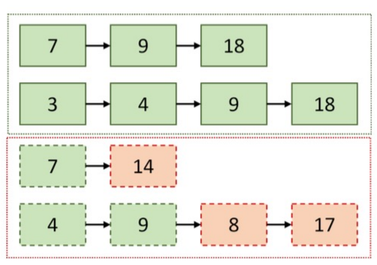


- First, use random-sampled prompts to generate diverse reasoning paths
- Then, use these reasoning paths to train a scoring verifier
- During inference time, use the verifier scores and do weighted-voting to get the final answer

Step-Aware Verifier



- First obtain data with step-level labels (figure on the right)
- Then, train a token-classification model as the step-aware verifier
- During inference time, use the step-aware verifier to score reasoning paths and do weighted-voting to get the final answer



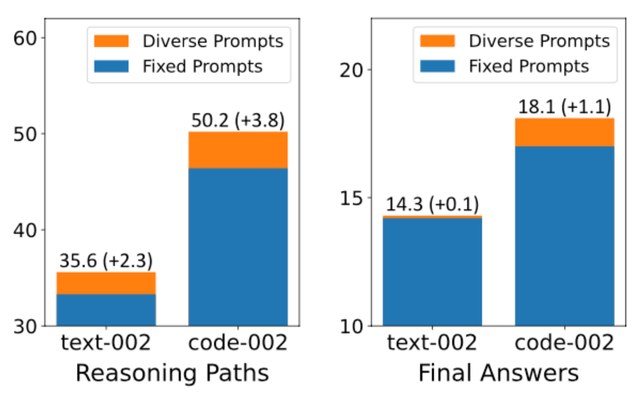
Using sub-paths of correct paths to label correct steps of wrong reasoning paths

Experiments

Method	GSM8K	AsDiv	MultiArith	SVAMP	SingleEq	CommonsenseQA	StrategyQA	CLUTRR
Previous SOTA (Fine-tuning)	57[a]	75.3[b]	60.5[c]	57.4[d]	32.5[e]	91.2[f]	73.9[g]	67.0 [h]
9–12 year olds (Cobbe et al., 2021)	60	-	-	-	-	-	-	-
LaMDA 137B:								
Greedy Decode	17.1	49.0	51.8	38.9	56.6	57.9	65.4	-
Self-Consistency	27.7	58.2	75.7	53.3	-	63.1	67.8	-
PaLM 540B:								
Greedy Decode	56.5	74.0	94.7	79.0	79.5	79.0	75.3	-
Self-Consistency	74.4	81.9	99.3	86.6	-	**80.7**	**81.6**	-
GPT-3 davinci (175B):								
Greedy Decode	8.7	31.4	31.4	21.2	38.2	48.2	59.2	33.6
Self-Consistency	18.9	52.8	68.6	44.6	59.6	57.4	65.6	42.5
DIVERSE	30.9 (+12.0)	57.6 (+4.8)	87.6 (+19.0)	46.9 (+2.3)	65.1 (+5.5)	75.0 (+17.6)	66.3 (+0.7)	92.5 (+50.0)
text-davinci-002:								
Greedy Decode	37.1	60.8	70.7	60.0	73.3	65.5	57.8	32.4
Self-Consistency	58.2	76.9	88.4	78.2	87.2	72.9	69.8	34.9
DIVERSE	70.2 (+12.0)	83.5 (+6.6)	96.4 (+8.0)	82.7 (+4.5)	86.5 (-0.7)	79.2 (+6.3)	74.8 (+5.0)	93.8 (+58.9)
code-davinci-002:								
Greedy Decode	55.3	75.5	88.8	70.5	87.5	73.4	72.0	32.9
Self-Consistency	76.7	86.2	98.6	85.8	93.7	77.3	77.6	35.6
DIVERSE	**82.3** (+5.6)	**88.7** (+1.5)	**99.8** (+1.2)	**87.0** (+1.2)	**94.9** (+1.2)	79.9 (+2.6)	78.6 (+1.0)	**95.9** (+60.3)

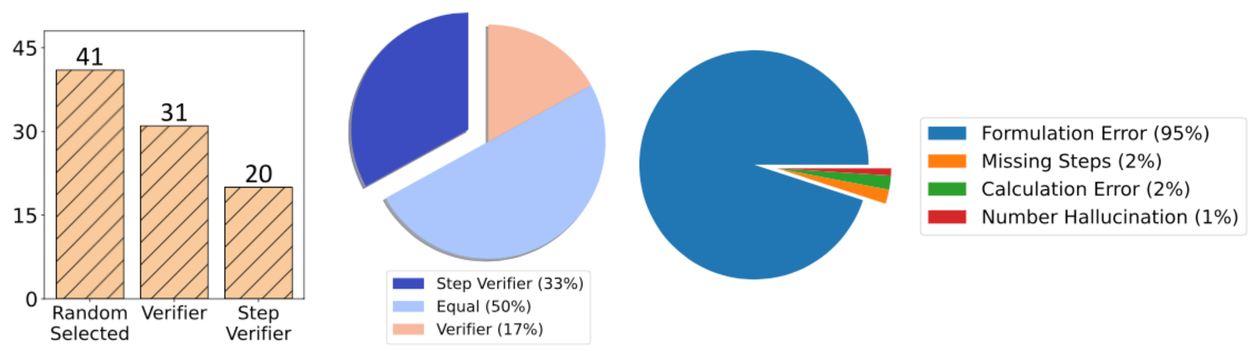
- **SoTA** on GSM8K, AsDiv, MultiArith, SVAMP, SingleEq, CLUTRR (June 7, 2022)
- Performance improved on 3 reasoning tasks (arithmetic, commonsense, inductive
- Outperforms PaLM 540B via GPT-3 code-davinci-002 model

Advantages Brought by Diversity



- Diverse prompts increase the diversity of GSM8K reasoning paths and their final answers.
- This diversity gain is also beneficial for the voting verifier since it augments the verifier's positive and negative samples.

Advantages of Step-Aware Verifier



- Step-aware verifier can further **eliminate redundant reasoning steps**
- The step-aware mechanism helps **produce more correct steps** before making mistakes
- Both verifiers help distinguish calculation errors

Case Study

Question	Self-Consistency	DIVERSE
[GSM8K] *Mark decides to buy packs of old magic cards and open them to sell. He buys 3 packs for $1800 each. He gets 1 card that is worth $4000 and another card worth $1000. There are 30 more cards worth an average of $50 each. How much money profit did he make?*	Mark bought 3 packs for $1800 each for a total of $1800 * 3 = 5400 (68.1%). He got 1 card worth $4000 and another card worth $1000 for a total of $4000 + $1000 = $5000 (72.1%). There were 30 more cards worth an average of $50 each for a total of $50 * 30 = 1500 (63.7%). Mark made a profit of $5000 + $1500 = 6500 (40.5%). The answer is 6500. **Answer:** 6500 (Wrong) **Score:** 62.3%	The cheaper cards are worth a total of 30 * 50 = 1500 (84.4%). So all the cards together are worth 1500 + 4000 + 1000 = 6500 (83.3%). The cost for the cards was 1800 * 3 = 5400 (80.1%). So he made a profit of 6500 - 5400 = 1100 (70.0%). The answer is 1100. **Answer:** 1100 (Correct) **Score:** 91.2%

- The step-level score given by the step-aware verifier shows where the sequence goes wrong

Conclusion

DIVERSE advances the reasoning capabilities in three aspects:

- **Diverse reasoning paths --** *"wisdom of the crowd"*
- **Voting verification --** *"reflective thinking"*
- **Step Correctness --** *"multistep thinking"*

DIVERSE can be applied on any LLMs regardless of the model architecture

Detailed analysis of the step-aware verifier

Limitations

- Our method need to be applied on LLMs like GPT-3 or PaLM
- As a common problem, the generated paths are not 100 percent faithful
- Human evaluations on steps may be replaced by better automatic metrics