Large Language Models Can Self-Improve

Jiaxin Huang[1*] **Shixiang Shane Gu**[2] **Le Hou**[2†] **Yuexin Wu**[2] **Xuezhi Wang**[2]
Hongkun Yu[2] **Jiawei Han**[1]
[1]University of Illinois at Urbana-Champaign [2]Google
[1]{jiaxinh3, hanj}@illinois.edu [2]{shanegu, lehou, crickwu, xuezhiw, hongkuny}@google.com

Venue : 2023 EMNLP

발제자 : 이다현 (hyundai@soongsil.ac.kr)

HUMANE Lab

2024-02-20

CC BY SA

Introduction

- LLM은 in-context few-shot learning으로 처음 본 task에서도 좋은 성능을 내게 됨
- 그러나 여전히 많은 양의 supervised data에 의존
- 본 연구에서는 Supervised data 없이도 LLM의 추론 능력을
- Self-Improve 하는 방법을 제안
- 본 연구의 목적:
  - Input data 만으로도 In-Domain과 Out-Of-Domain에서 좋은 성능을 내는 것

Introduction

- Self-consistency란?

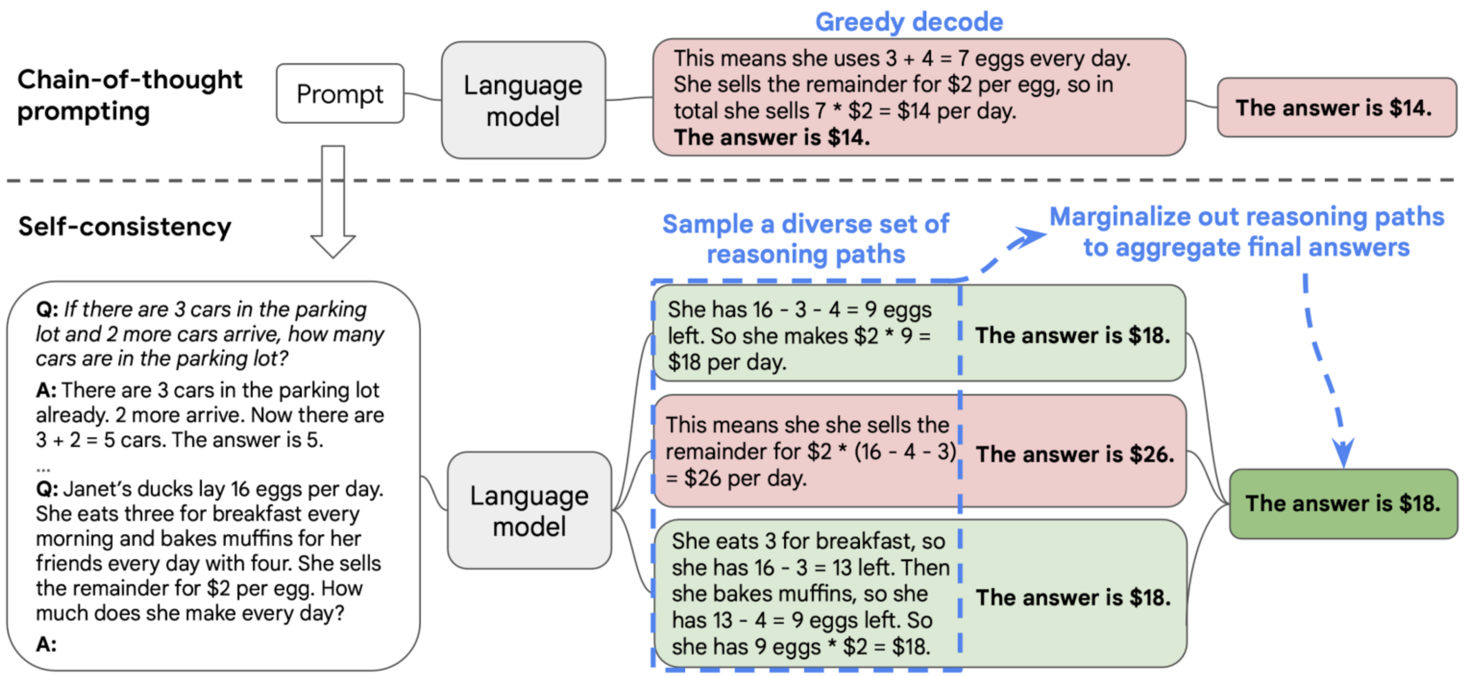


Image source: Wang et al., 2022, 'Self-Consistency Improves Chain of Thought Reasoning in Language Models,' Figure 2.

Method – Language Model Self Improved

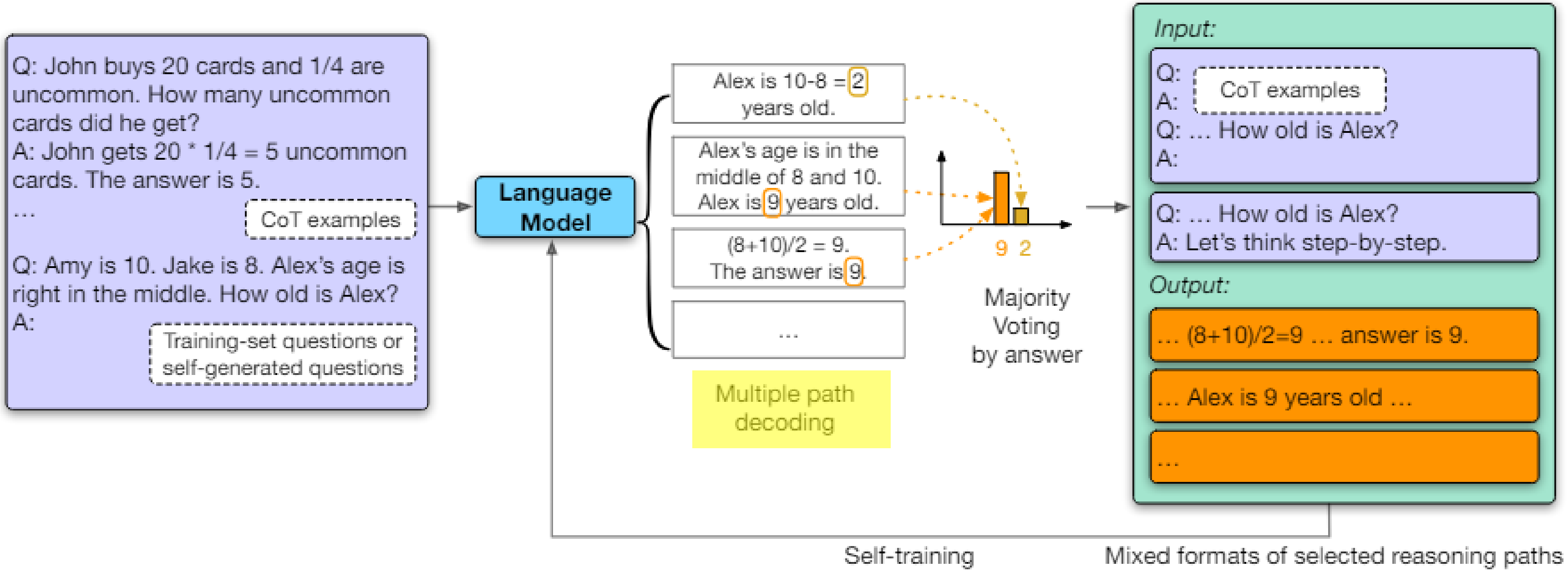
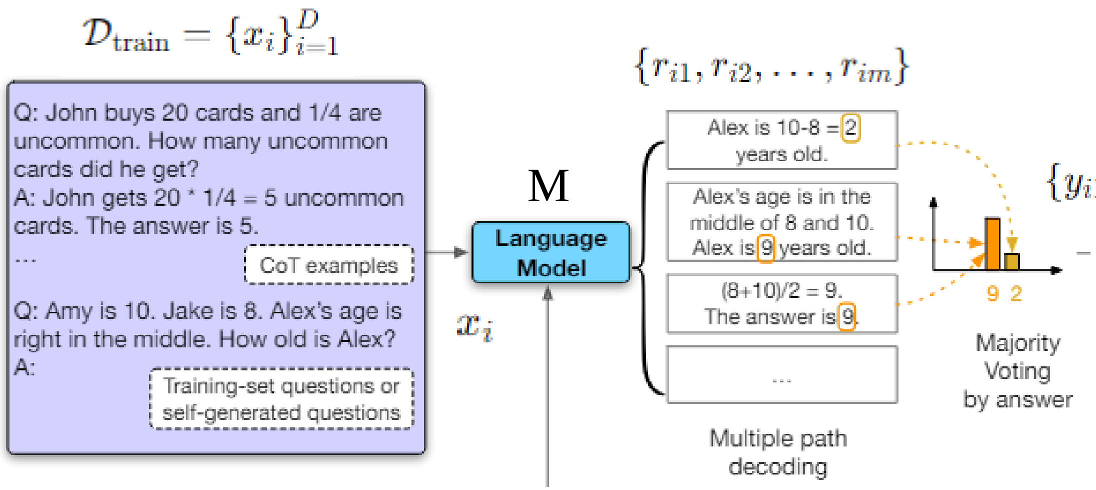


Figure 1: Overview of our method. With Chain-of-Thought (CoT) examples as demonstration (Wei et al., 2022c), the language model generates multiple CoT reasoning paths and answers (temperature $T > 0$) for each question. The most consistent answer is selected by majority voting (Wang et al., 2022c). The CoT reasoning paths that lead to the answer with the highest confidence are augmented by mixed formats, and are fed back to the model as the final training samples.

- LMSI는 in-context few-shot learning 과 CoT reasoning이 가능한 모델에만 적용할 수 있음

Generating and Filtering Multiple Reasoning Paths



$$\tilde{y}_i = \arg\max_{y_i} \sum_{k=1}^{m} \mathbb{I}(y_{ij} = y_{ik})$$

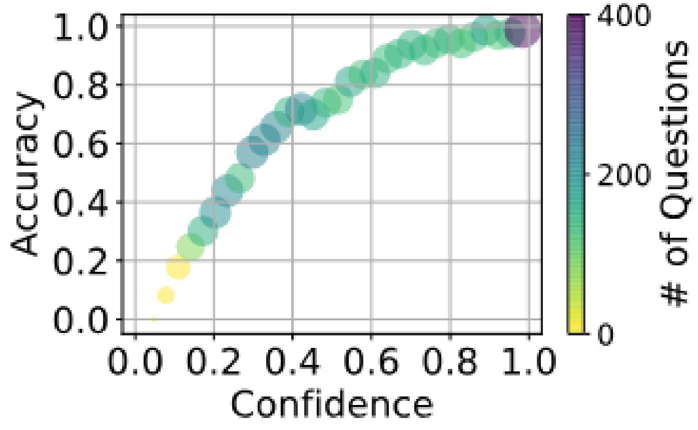


그림 2: GSM8K 훈련 세트의 각 question에 대해 self-generated된 CoT path에 대한 accuracy와 confidence의 관계

Question: Stefan goes to a restaurant with his family. They order an appetizer that costs \$10 and 4 entrees that are \$20 each. If they tip 20% of the total, what is the total amount of money that they spend?

Multiple Path Decoding:

Output 1: The appetizer costs \$10. The entrees cost \$20 each so in total 4 * \$20 = \$80. This means the total cost is \$10 + \$80 = \$90. They tip 20% of it, so the total amount they spend is \$90 * 1.2 = \$108. The answer is 108.

Output 2: The appetizer costs \$10 and the entrees are \$20 each. There are 4 entrees so the sum is \$20 * 4 = \$80. The waiter gets 20% of the total. 20% of \$80 is \$80 * .2 = \$16. The answer is \$80 + \$16 = \$96. *(Incorrect reasoning path)*

Output 3: The appetizer costs \$10. The entrees cost 4 * \$20 = \$80. The tip is 20% of the total, so it is 20% of the \$90 they have spent. The tip is 0.2 * 90 = \$18. The total they spent is \$90 + \$18 = \$108. The answer is 108.

$$\mathcal{D}^{\text{self-consistent}} = \{x_i, \tilde{r}_i\}$$

$$\tilde{r}_i = \{r_{ij} | 1 \le j \le m, y_{ij} = \tilde{y}_i\}$$

Training with Mixed Formats

Question: Amy is 10 years old. Jake is 8 years old. Alex's age is right in the middle. How old is Alex?
Selected Chain-of-Thought: Amy is 10 years old. Jake is 8 years old. Alex's age is in the middle of Amy and Jake, so Alex is (8 + 10) / 2 = 9 years old. The answer is 9.

Mixed-formats of training data:
Format 1: **Input**: *[CoT prompting examples]* + '\n' + *[Question]* + '\n' + 'A:'
Output: Amy is 10 years old. Jake is 8 years old. Alex's age is in the middle of Amy and Jake, so Alex is (8 + 10) / 2 = 9 years old. The answer is 9.

Format 2: **Input**: *[Standard prompting examples]* + '\n' + *[Question]* + '\n' + 'A:'
Output: The answer is 9.

Format 3: **Input**: *[Question]* + '\n' + 'A: Let's think step by step.'
Output: Amy is 10 years old. Jake is 8 years old. Alex's age is in the middle of Amy and Jake, so Alex is (8 + 10) / 2 = 9 years old. The answer is 9.

Format 4: **Input**: *[Question]* + '\n' + 'A:'
Output: The answer is 9.

	Examples	Chain of thought
Few-shot CoT	O	O
Few-shot	O	X
Zero-shot CoT	X	O
Zero shot	X	X

Self-Training

사전학습된 언어모델 M을 미세조정 하는데 사용

Generating Questions and Prompts

- Training question이나 human-curated CoT 프롬프트가 제한된 몇몇 경우에서 제안된 방법은 제한적
- 사람의 개입을 최소화하는 방법 제시
  - **Question Generation**
    - 무작위로 뽑은 question을 무작위 순서로 연결하여 프롬프트 생성
    - 프롬프트를 입력하여 LLM이 새로운 질문을 생성하도록 유도
    - Self-Consistency 적용하여 선별
  - **Prompt Generation**
    - Step-by-Step 방법 (Kojima et al., 2022)을 사용
    - 답변을 “A: Let’s think step by step.”으로 시작하여 언어모델이 추론 경로를 생성하도록 유도

Experimental Setup - Datasets

- Arithmetic reasoning (산술 추론)
  - GSM8K
  - DROP

- Commonsense reasoning (상식 추론)
  - OpenBookQA
  - AI2Reasoning Challenge

- Natural Language Inference (자연어 추론)
  - Adversarial NLI

Experimental Setup - Models

- PaLM 540B
    - Autoregressive 트랜스포머 기반 언어 모델
- 각 질문에 대한 추론 경로 32개 생성
- 샘플링 온도 T = 0.7 (Wang et al., 2022c에서 제안)
- 하이퍼파라미터 세팅
    - 10k step 학습
    - Learning rate: 5e-5
    - Batch size: 32
- 디코딩 단계의 최대 수: 256

Experiements

1. 각 데이터셋(task)에 제안한 방법을 적용
2. 데이터셋에서 생성된 결과를 통합하여 하나의 모델을 미세 조정
    - Wei et al. (2021)과 같이
      미확인 데이터셋에 대한 모델의 일반화 능력 확인
3. 입력 질문과 Few-shot 프롬프트 생성에 대한 실험
4. 모델 크기와 하이퍼파라미터 영향 실험

Results

Table 3: Accuracy results on six reasoning benchmarks with or without **LMSI** using different prompting method.

Prompting Method	w. or w/o **LMSI**	**GSM8K**	**DROP**	**ARC-c**	**OpenBookQA**	**ANLI-A2**	**ANLI-A3**
Standard-Prompting	w/o **LMSI**	17.9	60.0	87.1	84.4	55.8	55.8
	w. **LMSI**	32.2 (+14.3)	71.7 (+11.7)	87.2 (+0.1)	92.0 (+7.6)	64.8 (+9.0)	66.9 (+11.1)
CoT-Prompting	w/o **LMSI**	56.5	70.6	85.2	86.4	58.9	60.6
	w. **LMSI**	73.5 (+17.0)	76.2 (+5.6)	88.3 (+3.1)	93.0 (+6.6)	65.3 (+6.4)	67.3 (+6.7)
Self-Consistency	w/o **LMSI**	74.4	78.2	88.7	90.0	64.5	63.4
	w. **LMSI**	**82.1** (+7.7)	**83.0** (+4.8)	**89.8** (+1.1)	**94.4** (+4.4)	**66.5** (+2.0)	**67.9** (+4.5)

- 여섯개의 벤치마크 데이터셋에서 각 프롬프팅 방법에 대한 LMSI 적용 전후를 비교

Results - Multi-task self-training for unseen tasks

Table 4: Comparison of CoT-prompting accuracy results on six Out-Of-Domain benchmarks with or without training on six In-Domain (GSM8K, DROP, ARC-c, OpenBookQA, ANLI-A2, ANLI-A3) training-set questions.

	Self-training data	AQUA	SVAMP	StrategyQA	ANLI-A1	RTE	MNLI-M/MM
w/o **LMSI**	-	35.8	79.0	75.3	68.8	79.1	72.0/74.0
w. **LMSI**	GSM8K + DROP + ...	39.0 (+3.2)	82.8 (+3.8)	77.8 (+2.5)	79.2 (+10.4)	80.1 (+1.0)	81.8/82.2 (+9.8/+8.2)

- LMSI의 일반화 성능을 입증
  - 여섯개 데이터셋(in-domain task)의 training set questions를 혼합하여 self-training 수행
  - 동일한 모델 체크포인트를 사용하여 Out-of-Domain 작업에 대한 평가를 진행

Results - Importance of training with augmented formats

Table 5: Ablation study: **LMSI** with different combinations of training format on GSM8K dataset.

	Results on **GSM8K** Std. Prompting	CoT Prompting
w/o **LMSI**	17.9	56.5
LMSI w/o CoT formats	23.6 (+5.7)	61.6 (+5.1)
LMSI only few-shot CoT	29.2 (+11.3)	69.4 (+12.9)
LMSI w/ CoT formats	32.2 (+14.3)	73.5 (+17.0)

- 증강된 형식으로 언어 모델을 훈련하는 것의 중요성을 입증

Results - Pushing the limit of self-improvements

- **Self-Generating Questions**

Table 6: Accuracy on GSM8K test set after self-training on different question sets. Results are shown for both CoT-Prompting (CoT) and Self-Consistency (SC).

	Questions used for Self-Training	GSM8K CoT	GSM8K SC
w/o **LMSI**	-	56.5	74.4
w. **LMSI**	Generated	66.2 (+9.7)	78.1 (+3.7)
w. **LMSI**	Training-set	73.5 (+17.0)	82.1 (+7.7)

- GSM8K 데이터셋에서 10개의 질문을 few-shot 샘플로 선정
  - 언어모델을 사용해서 질문을 생성

- **Self-Generating Few-Shot CoT Prompts**

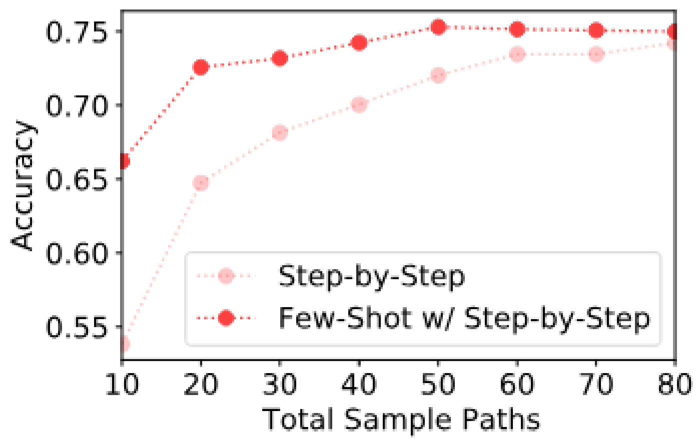


Figure 3: Accuracy results on GSM8K test set using 540B model with multi-path sampling and self-consistency (Wang et al., 2022c). "Step-by-Step" is the baseline performance of Kojima et al. (2022) plus self-consistency (Wang et al., 2022c), while our "Few-Shot w/ Step-by-Step" uses exemplers self-generated from Step-by-Step (greedy decoding) for few-shot prompting the LLM.

Results - Distillation to smaller models

Table 7: Distillation from 540B model to small models. We see that distilled smaller models outperform models that are one-tier larger.

	Results on **GSM8K**		
	8 billion	62 billion	540 billion
w/o **LMSI**	5.0	29.7	56.5
Distilled from **LMSI**	33.4 (+28.4)	57.4 (+27.7)	-

- 지식이 더 작은 모델로 distill(증류) 될 수 있는지 탐구
  - PaLM 540B 모델로 생성된 동일한 train sample set를 사용
  - 더 작은 크기의 모델을 Fine-tuning 함

Results - Hyperparameter Studies

- **Sampling Temperature after Self-Improvement**

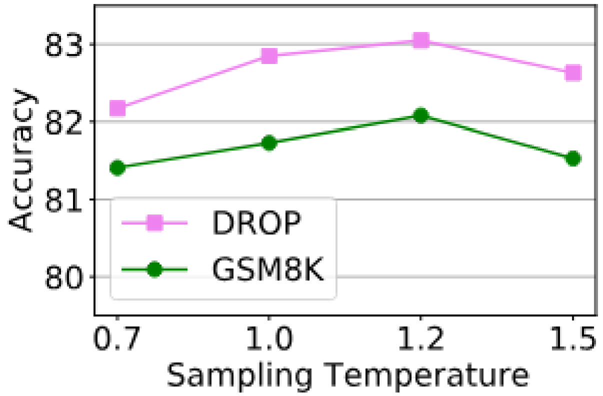


Figure 4: Accuracy results of **LMSI** on GSM8K and DROP test set when different sampling temperatures are applied for Self-Consistency.

- Temperature
  - 모델이 텍스트를 생성할 때 예측 분포의 불확실성을 조정하는 매개변수

- **Number of Sampled Reasoning Paths**

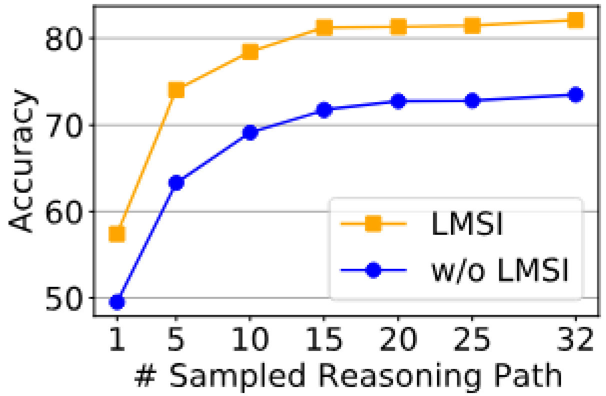


Figure 5: Accuracy results with or without **LMSI** on GSM8K test set using different numbers of sampled reasoning path for Self-Consistency.

Conclusions

- LLM이 입력 질문만으로도 LLM이 자체적으로 생성한 label을 통해 추론 능력 향상이 가능함을 입증
- LMSI의 효과성
  - 540B PaLM 모델 실험에서 LMSI가 여러 데이터셋에서 점수를 향상시킴을 확인
- LLM이 자체 생성한 질문과 few-shot CoT 프롬프트를 사용해서도 성능 향상이 가능함을 확인

Open question

- LLM의 기본 성능이 매우 낮은 domain에서도 자체 생성한 질문과 few-shot CoT 프롬프트를 사용해 성능 향상이 가능할까?