ReAct: Synergizing **Re**asoning and **Act**ing in Language Models

Berk Atil

04/03/2023

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Outline

- Motivation
- Methodology
- Results on Knowledge Intensive Reasoning
- Results on Decision Making
- Human in the loop
- Conclusion

Motivation and Background

- **Prompting:** Basically, we embed the task description that we want from the model to solve into the input.
- **Prompt Engineering:** The focus is more on the prompts that we provide to the model, trying to figure out a template/structure of a prompt that performs best.
- This can be used in both zero-shot and few-shot setting.
- [1] is one of the earliest work on prompting where they show that few-shot learning can provide good performance instead of fine-tuning models.

[1] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Motivation

- Humans combine verbal reasoning with actions.
- Synergy between reasoning and acting
- Emergent capability of language models [1].
  - Only internal knowledge, no update on knowledge
  - Fact Hallucination
- LMs can be used for planning actions in interactive environments [2]
  - No reasoning, no action taking

[1] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*.
[2] Brohan, A., Chebotar, Y., Finn, C., Hausman, K., Herzog, A., Ho, D., ... & Fu, C. K. (2023, March). Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on Robot Learning* (pp. 287-318). PMLR.

(1) Hotspot QA

Question: Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

(1a) Standard

Answer: iPod

(1b) CoT (Reason Only)

Thought: Let's think step by step. Apple Remote was originally designed to interact with Apple TV. Apple TV can be controlled by iPhone, iPad, and iPod Touch. So the answer is iPhone, iPad, and iPod Touch.

Answer: iPhone, iPad, iPod Touch

(1c) Act-Only

Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control …

Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: …

Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software …

Act 4: Finish[yes]

Why Not Combining These Two?

- Models can reason to act to create and adjust plans
- Also they can act to reason by interacting with external environments

(1d) ReAct (Reason + Act)

Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with.
Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple … originally designed to control the Front Row media center program …

Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.
Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports',' Front Row (software) ', …]

Thought 3: Front Row is not found. I need to search Front Row (software) .
Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software …

Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.
Act 4: Finish[keyboard function keys]

ReAct

- Prompting based approach with PaLM [1]
- Enlarge the Action Space with Language Space
- Sequence of “Thought”, “Action”, “Observation”
- Easy to design
- Flexible
- Human aligned and controllable

[1] Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., ... & Fiedel, N. (2022). Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.

Knowledge Intensive Reasoning

- Multi-hop Q&A, 6-shot
- Fact Verification, 3-shot
- 3 actions: search, lookup, finish
- Multiple thought-action-observation steps
- ReAct provides grounded problem solving but chain-of–thought provides more accurate reasoning
  - Incorporate both and let model switch from one to the other.

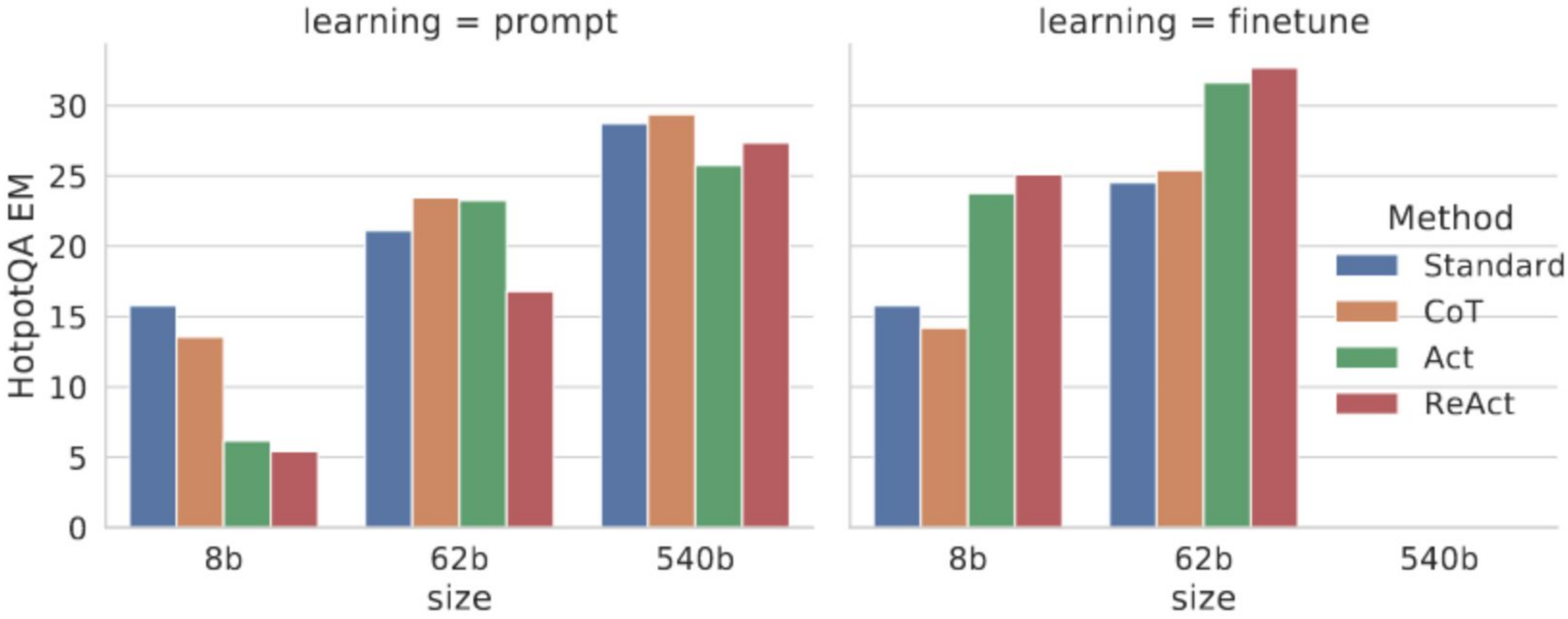
Results

- Reasoning is important to guide acting
- ReAct vs CoT
- ReAct + CoT-SC [1] is the best

Prompt Method[a]	HotpotQA (EM)	Fever (Acc)
Standard	28.7	57.1
CoT (Wei et al., 2022)	29.4	56.3
CoT-SC (Wang et al., 2022a)	33.4	60.4
Act	25.7	58.9
ReAct	27.4	60.9
CoT-SC → ReAct	34.2	**64.6**
ReAct→ CoT-SC	**35.1**	62.0
Supervised SoTA[b]	67.5	89.5

[1] Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., & Zhou, D. (2022). Rationale-augmented ensembles in language models. *arXiv preprint arXiv:2207.00747*.

Finetuning Results



Decision Making

- ALFWorld [1], synthetic text-based game
  - 6 types of tasks that an agent needs to achieve a high-level goal.
  - Sparse thoughts/reasonings in prompts.
- WebShop [2], online shopping website environment.
  - Based on user instructions, it should buy a product.

[1] Shridhar, M., Yuan, X., Côté, M. A., Bisk, Y., Trischler, A., & Hausknecht, M. (2020). Alfworld: Aligning text and embodied environments for interactive learning. *arXiv preprint arXiv:2010.03768*.
[2] Yao, S., Chen, H., Yang, J., & Narasimhan, K. (2022). Webshop: Towards scalable real-world web interaction with grounded language agents. *arXiv preprint arXiv:2207.01206*.

Results on ALFWorld

Method	Pick	Clean	Heat	Cool	Look	Pick 2	All
`Act` (best of 6)	88	42	74	67	72	**41**	45
`ReAct` (avg)	65	39	83	76	55	24	57
`ReAct` (best of 6)	**92**	58	**96**	86	**78**	**41**	**71**
`ReAct-IM` (avg)	55	59	60	55	23	24	48
`ReAct-IM` (best of 6)	62	**68**	87	57	39	33	53
BUTLER$_g$ (best of 8)	33	26	70	76	17	12	22
BUTLER (best of 8)	46	39	74	**100**	22	24	37

- It depends highly on the prompts
- Sparse reasoning helps.

Results on WebShop

Method	Avg Score	Success
`Act`	62.3	30.1
`ReAct`	**66.6**	**40.0**
IL	59.9	29.1
IL+RL	62.4	28.7
Human Expert	82.1	59.6

Human in the loop correction

- Edit and correct ReAct's false reasonings

AlfWorld

You are in the middle of a room. Looking quickly around you, you see a armchair 1, a cabinet 4, … a cabinet 1, a drawer 5, …, a drawer 1, a dresser 1, a garbagecan 1, a safe 1, a shelf 12, …, a shelf 1, a sidetable 1, and a sofa 1.
Your task is to: put two keychain in safe.

(a) ReAct

......
Obs 16: You open the drawer 4. The drawer 4 is open. In it, you see a keychain 3, and a watch 2.
Act 17: Think [Now I find the first keychain (3). Next, I need to take it. I can find the second keychain (2) later in drawer 4.]
Act 18: take keychain 3 from drawer 4
Obs 18: You pick up the keychain 3 from the drawer 4.
......
Act 22: put keychain 3 in/on safe 1
Obs 22: You put the keychain 3 in/on the safe 1.
Act 23: Think [Now I put the first keychain in safe. Next, I need to find the second keychain. I can directly go to drawer 4.]
Act 24: go to drawer 4
Obs 24: The drawer 4 is open. In it, you see a watch 2.
Act 25: Think [Now I find the second keychain (2). Next, I need to take it.]
Act 26: take keychain 2 from drawer 4
Obs 26: Nothing happens.

(b) ReAct (Human Thought edit)

......
Obs 16: You open the drawer 4. The drawer 4 is open. In it, you see a keychain 3, and a watch 2.
Act 17: Think [Now I find the first keychain (3). Next, I need to take it. ~~I can find the second keychain (2) later in drawer 4.~~]
Act 18: take keychain 3 from drawer 4
Obs 18: You pick up the keychain 3 from the drawer 4.
......
Act 22: put keychain 3 in/on safe 1
Obs 22: You put the keychain 3 in/on the safe 1.
Act 23: Think [Now I put the first keychain in safe. Next, I need to find the second keychain, which is more likely to appear in dresser (1), garbagecan (1), safe (1), sidetable (1), sofa (1), shelf (1-12).]
Act 24: go to dresser 1
Obs 24: On the dresser 1, you see a box 1, a creditcard 1, a keychain 2, a keychain 1, a laptop 1...
Act 25: Think [Now I find the second keychain (2). Next, I need to take it.]
Act 26: take keychain 2 from dresser 1
Obs 26: You pick up the keychain 2 from the dresser 1.
Act 27: Think [Now I take the second keychain (2). Next, I need to put it in/on safe 1.]
Act 28: go to safe 1
Obs 28: The safe 1 is open. In it, you see a keychain 3, a statue 2, and a watch 3.
Act 29: put keychain 2 in/on safe 1
Obs 29: You put the keychain 2 in/on the safe 1.

Conclusion, Limitations and Future Work

- Reasoning and action helps models reach to better conclusions.
- There is still room for improvement, especially for knowledge intensive reasoning tasks it is behind the supervised methods
- Prompt design affects the performance
- When the action space is large the size of the incontext learning grows quickly.
- More high quality labelled data for finetuning
- Incorporation with Toolformer [1], a language model that can interact with more tools such as calculator, calendar, translator etc.

[1] Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., ... & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. *arXiv preprint arXiv:2302.04761*.

Thanks!

- Any Questions?