REAC T: SYNERGIZING REASONING AND ACTING IN LANGUAGE MODELS

Problem:

CoT: no access to the external world ->Can not reason reactively or update its knowledge -> hallucination and error propagation

Action: do not employ language models to reason abstractly about high-level goals or maintain a working memory

Methodology:

at ∈ A

policy π(at|ct)

ct = (o1, a1, · · · , ot−1, at−1, ot)

ct → at (highly diffcult)

Through reasoning:

Aˆ = A ∪ L, aˆt ∈ L, L is unlimited

ct+1 = (ct, aˆt)

Advantage:

* Intuitive and easy to design
* General and flexible
* Performant and robust
* Human aligned and controllable

Experiment:

CoT + Search Wikipedia

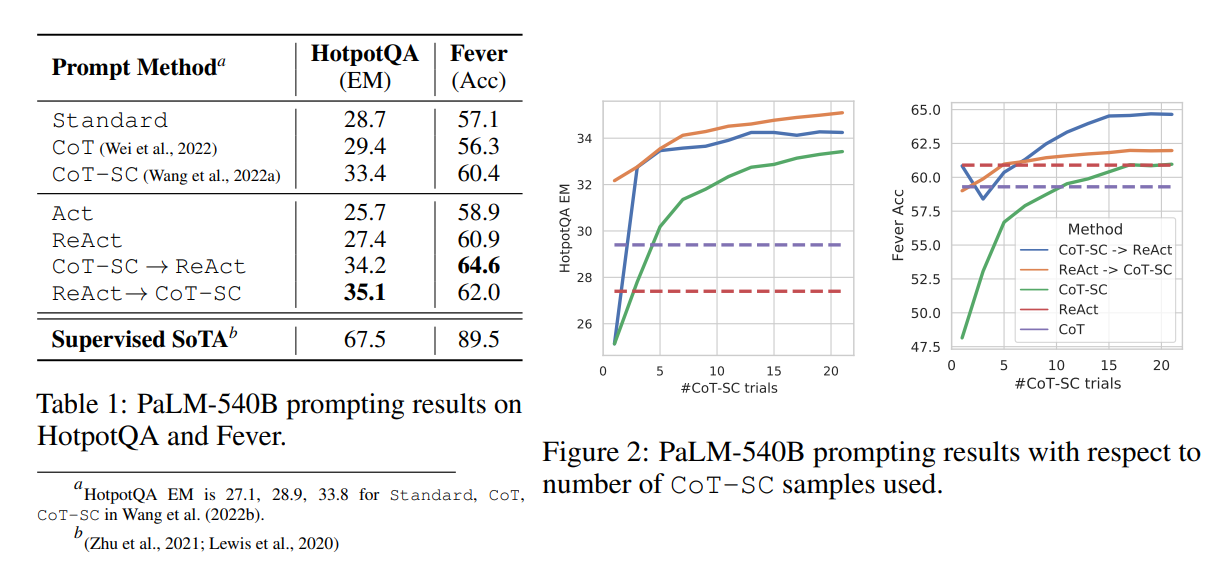
Action Space:

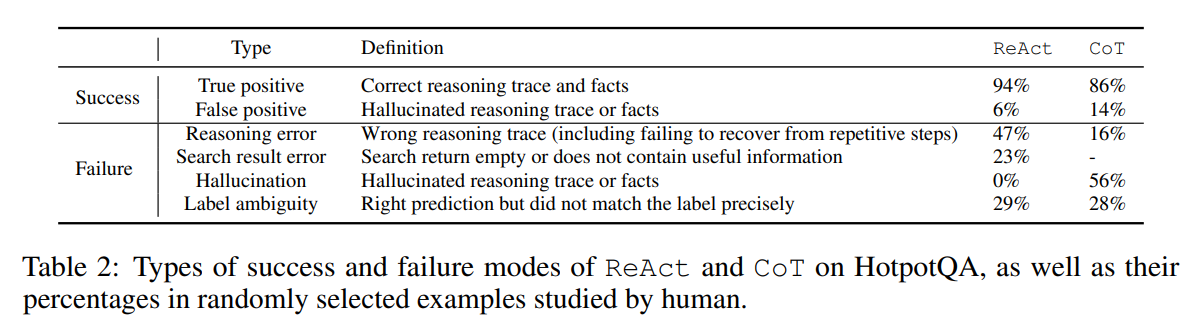
1. search[entity] returns the first 5 sentences from the corresponding entity wiki page
2. lookup[string] Ctrl+F
3. finish[answer]

manually compose ReAct-format trajectories to use as few-shot exemplars in the prompts

4 Benchmarks: (HotpotQA), (Fever), (ALFWorld), (WebShop).

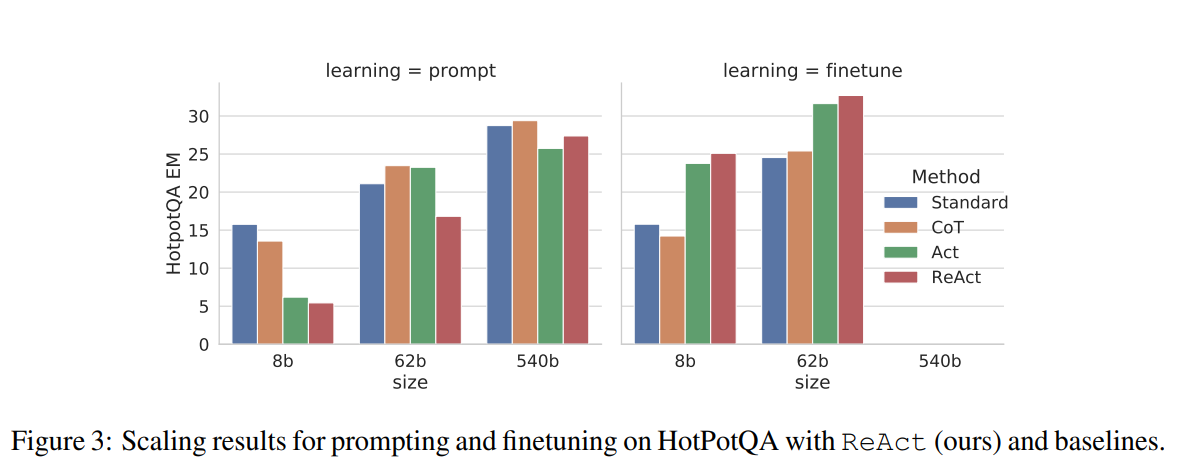
3 baselines: Standard, Chain-of-Thought prompting, Acting-only prompt

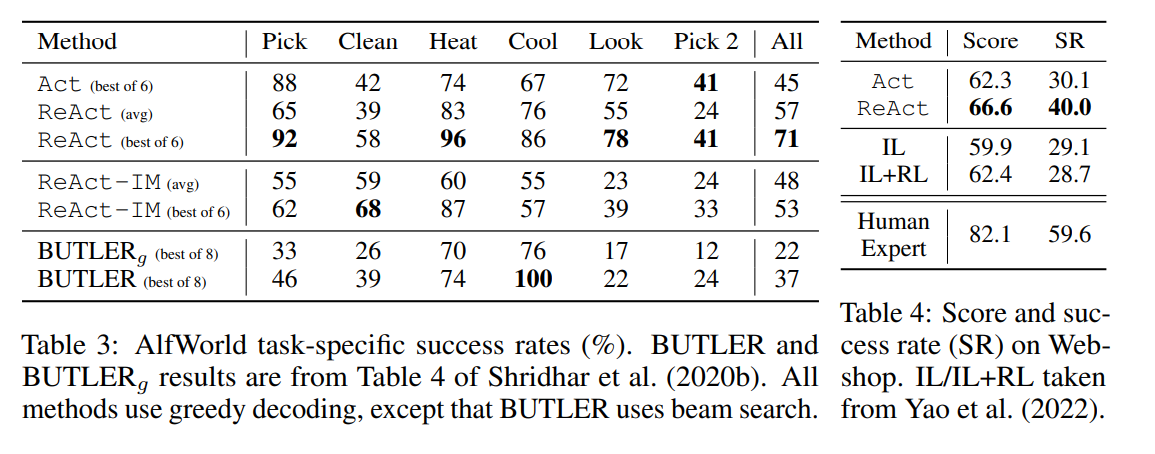




Concretely, on question answering (HotpotQA) and fact verification (Fever), ReAct overcomes prevalent issues of hallucination and error propagation in chain-of-thought reasoning by interacting with a simple Wikipedia API.

on two interactive decision making benchmarks (ALFWorld and WebShop), ReAct outperforms imitation and reinforcement learning methods by an absolute success rate of 34% and 10% respectively, while being prompted with only one or two in-context examples.





Comparison:

* Hallucination is a serious problem for CoT
* While interleaving reasoning, action and observation steps improves ReAct’s groundedness and trustworthiness, such a structural constraint also reduces its flexibility in formulating reasoning steps
* For ReAct, successfully retrieving informative knowledge via search is critical.
* ReAct + CoT-SC perform best for prompting LLMs
* ReAct performs best for fine-tuning