Hey AI, Can You Solve Complex Tasks by Talking to Agents?

Tushar Khot, Kyle Richardson, Daniel Khashabi, Ashish Sabharwal

Motivation

Can you buy me the book series that has the kid with that lightning scar?

Hey Google, Which book series has a kid with a lightning scar?

Hey Alexa, Buy Harry Potter and the Sorcerer’s Stone

Summary

New Task: Learning to Talk to Agents to Solve Complex Tasks

1. Solve a complex task by breaking it down into agent’s capabilities
2. Interact with agents in their expected and natural language

Why?

- Green AI: Reuse existing expensive and even proprietary models
- Better Long-Term Bet: No need to learn every task from scratch
- Interpretability: Naturally modular and interpretable systems
- Technical Challenge: Search for solutions by interacting with NL agents

New Dataset: CommaQA: Communicating with Agents for QA

- Synthetic Multi-hop QA Dataset solvable using agents
- Challenging for current black-box models and task baselines

https://github.com/allenai/CommaQA

Task: Learning to Talk to Agents to Solve Complex Tasks

Given:

Agents

Examples of Valid Inputs

- What is the list of books in LOTR series
- Which movies are in the Fast & Furious series
- What is the list of books in the Harry Potter series?

To Do:

Learn to solve the complex task by communicating with agents

Should NOT rely on:

Training Data of the Agents

Agents Internal Knowledge

Conversation Supervision

Dataset: CommaQA -- Communicating with Agents for QA

CommaQA-E

Inspired by HotpotQA (Yang et al. ‘18), MusicQuiz (Trivedi et al. ‘21)

CommaQA-I

Inspired by OpenBookQA (Mihaylov et al. ‘18), StrategyQA (Geva et al. ‘21)

CommaQA-N

Inspired by DROID (Dua et al. ‘19)

Related Work

Multi-hop QA Datasets: These datasets (Khashabi et al., 2018; Mihaylov et al., 2018; Yang et al., 2018; Dua et al., 2019; Khot et al., 2020; Geva et al., 2021) can be potentially solved by composition of single-hop models but

- Single-hop shortcuts incentivize non-compositional models (Min et al., 2019a; Trivedi et al., 2020)
- Lack reliable agents to solve single-hop sub-tasks (e.g. list answer QA) (Khot et al., 2021)

Question Decomposition: These approaches solve QA by decomposing complex question but,

- Current approaches generally limited to one (Talmor and Berant, 2018; Min et al., 2019b; Perez et al., 2020)
- QA model
- Many questions are out-of-scope due to lack of agents (Khot et al., 2021)
- Rely on human annotation of decomposition (Talmor and Berant, 2018; Min et al., 2019b)

Text-Based Games: Also require solving tasks by interacting with agents (often the game environment) but focus on different class of problems with different assumptions on agent’s language (Prian et al., 2019; 2020; Haakkerfelt et al., 2020; Ammanabrolu et al., 2021; Jansen, 2021)

Results

Unsolved using the current baselines that talk to agents

Black-box models struggle even when given access to the agent’s private knowledge

But solvable by training on conversation supervision (oracle upper bound)