**Composite Task-Completion Dialogue Policy Learning via Hierarchical Deep Reinforcement Learning**

Building a dialogue agent to fulfill complex tasks, such as travel planning, is challenging because the agent has to learn to collectively complete multiple subtasks.

-> address this challenge by formulating the task in the mathematical framework of options over **MDPs**, and **proposing a hierarchical deep reinforcement learning** approach to learning a dialogue manager.

**The dialogue manager consists of**

1) a top-level dialogue policy that selects among subtasks or options

2)a low-level dialogue policy that selects primitive actions to complete the subtask given by the top-level policy

3)a global state tracker that helps ensure all cross-subtask constraints be satisfied.

**Conversation as Reinforcement Learning**

User input->Language understanding->Dialogue Manager->Collect rewards->Optimize->Dialogue Manager->Language(response) generation->Respose

**A user simulator for RL and evaluation**

-Robustness : automatic action selection based on uncertainty by RL

-Flexibility : allow user-initiated behavors

-Reproducibility : a R&D setting that allows consistent comparisons of copeting methods

**Composite task completion bot with Hierarchical RL**

-Reward sparsity : takes many turns to fulfill a task

-Cross-subtask slot constraints

-Coherent conversation

(Travel Assistant에서 Book Flight, Book Hotel, Reserve Restaurant가 subtask)

**Exploiting structural info for efficient exploration**

-Internal Critic gives intrinsic reward(How likely a subtask is completed)

-Encourage agent to complete a subtask before moving to another subtask

-Better user experience : less freqency switch between subtasks

User simulators

1)Type A

-Have hard constraints for the flight and hotel(single value for all informed slots)

-No preference on which subtask to accomplish first

2)Type B

-At least one of informed slots in book-flight-ticket have multiple values

-Prefer to start with book-flight-ticket

3)Type C

-At least one of informed slots in reserve-hotel have multiple values

-Prefer to start with reserve-hotel

**Human Evaluations**

-HRL is most successful

-Rule based System은 자주 fail

\*State : User input+context

\*Action : Dialog act + slot value

\*Reward : Task success rate