Automated Negotiating Agent using Adaptive Q-learning

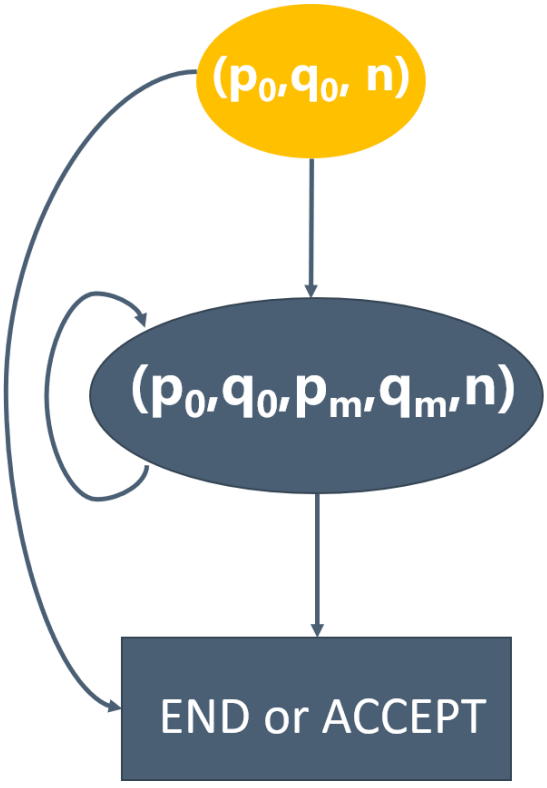
By LIZONGCAN

1. Agent description:

I decided to use Q-learning algorithm - a well-known and widely used reinforcement learning algorithm, as the base. Since the negotiation problem in the One-shot track can be modeled as an infinite/finite MDP, the q-learning algorithm should be able to converge to an optimal policy after many rounds of simulation. The code of q-learning algorithm came from an agent participated in the One-shot competition last year named Qlagent. For better improvement, I removed some of the extra strategies and kept the original q-learning strategy. As for the improvement, I implemented a strategy that uses q-learning as base. During the decision making process, there will be another significant party inside the agent called action advisory module. The module will be detailly introduced in the rest part of this report.

2. Algorithm in detail

2.1 States setting.



This agent follows the states setting of the original Qlagent.

The states inside this agent transit from initial states to intermediate states, and ends at the terminal states.

Initial state: (p_0, q_0, n) is used when the system called the agent to respond to an offer or to propose an offer. p_0 : the price of last offer received. q_0: the quantity of the last offer received. n: agent's need.

Intermediate states: (p_0, q_0, p_m, q_m, n) except for the offer information the agent received,

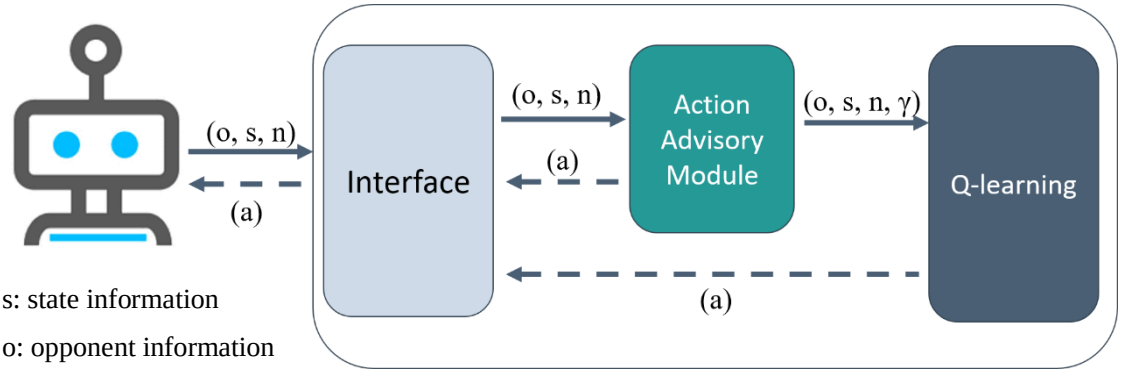
p_m, q_m : the offer information the agent previously offered.

Terminal state: END or ACCEPT.

For a simpler implementation, during the decision-making process the price and quantity space are assumed as finite, which means the agent selects actions to execute among finite options.

2.2 Adaptive Q-learning

The structure of my adaptive q-learning agent is described as below.



s: state information

o: opponent information

n: negotiation details

a: action

γ: reward bias

The core algorithm communicates with the opponent agent through an interface, while the action advisory module stands between the interface and q-learning algorithm to help the agent make better decision.

1) Interface:

This part does nothing significant but processing some raw information into something that can be easily used by the rest part of the agent, e.g. extracting the information from the offers and storing it inside or judging if the opponent agent is rational.

2) The details of Action advisory module:

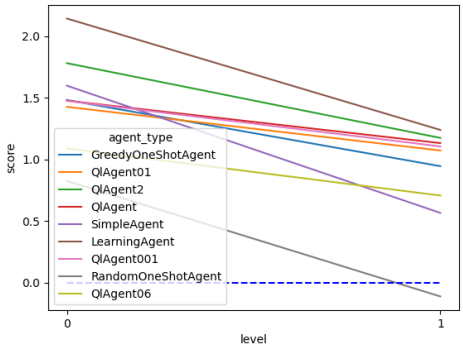
This module is the novel part which differentiates my adaptive q-learning agent from the original one. While the agent is having difficulty making decisions, e.g. the balance between exploration and exploitation, the action advisory module guides the q-learning algorithm by advising reward bias. For the simplest case, the agent may fall into continuous failures during the negotiation with a specific agent (this problem was proved to be real in my simulations). The action advisory module will start to work and affect the q-learning algorithm by giving reward bias.

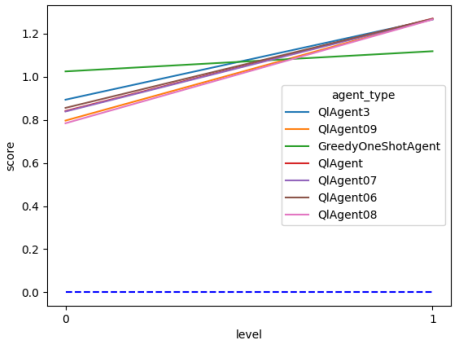
3) Q-learning algorithm.:

The q-learning part maintains different q tables for different agent, which allows it make opponent-based decisions.

3. Evaluation:

Adaptive q-learning agent is put into an environment to compete with other q-learning agent with different learning strategies, various ε-greedy strategy and various discounting factors.





In most cases, the adaptive q-learning agent was more stable and out-performed other q-learning agents.

4. Possible improvements to be made:

4.1 The rationality and irrationality.

The interface part judges the agent by telling it is rational or irrational. However, the judging condition is not fully studies and tested. Since the agents' actions are quite complex and there may be more improvements can be made for the interface.

4.2 Meta-learning techniques.

Implementing this module into a meta-learning algorithm is a better way to make q-learning adaptive. I believe there are already many researchers doing this so it is a possible and reasonable improvement.