Comparable Agents

yotamitai

July 2020

1 Introduction

An Reinforcement Learning agent is defined as policy π mapping states to actions. There are numerous methods as to how a policy should be constructed, be it through classical RL methods such as Policy Iteration or via an abstract learning of state features and weights as is the case in Deep RL. Either way, the outcome must be the same, such that the mapping be obtained.

2 Comparison Criteria

In order to compare agents to one another we require means to evaluate the difference between them.

2.1 Agent Disagreement

One such method, which is commonly used in the classical RL settings, is action comparison per state between two policies. The Policy Iteration algorithm utilizes this approach after each iteration to asses if the optimal policy has been reached. Agents (policies) are then deemed similar solely by their choice of actions in each state with no regard to other defining features such as the values assigned for each state.

In other words, any state s for which different agents choose different actions we will denote as a disagreement state. We will use these states to analyse and describe how and why these agents differ from one another in their performance. formally:

Definition 1 (Disagreement State) *Given n agents $\pi_1, \pi_2..., \pi_n$ and a state s. Define s_D as a disagreement state iff*

$$\exists i, j \; s.t. \; \pi_i(s) \neq \pi_j(s)$$

For a compact MDP for which every state may be computed, this definition could suffice. Alas for more complex settings where the state space is vast or continuos it is not feasible to compare all states and so most approaches use value functions in order to represent and asses a given state. These output an

approximation which hopefully captures the underlying dynamics of the task sufficiently enough as navigate our agent towards the desired direction.

Different agents in this setting may have different evaluations for state features. Therefore, given a state s each agent evaluates differently which action would lead him to a more promising next-state. If different actions are chosen we can flag this state as a disagreement state and try to analyse what parts of the feature evaluations is responsible for the disagreement in order to describe the differences between the agents.
On the other hand, if no disagreement has been found in the states we have observed, can we rest assured that these agents are truly similar?
Let us define a term for agents who are sufficiently similar:

Definition 2 (Sufficiently Similar) *Given two agents π_1, π_2, we denote the agents as sufficiently similar iff:*

$$\forall s,\ \pi_1(s) = \pi_2(s)$$

In other words, if two agents differ in their feature evaluations but still agree on the action to initiate for each state s, we will say that they are sufficiently similar.

Unfortunately, once again due the complexity of the state space, the above relation between agents is not feasible to attain. Therefore we require further means of identifying similarity (or disagreement) between agents.

2.2 Agent Confidence

Agent confidence can be described as how sure an agent is that the action he chose is the optimal one to initiate in the current state. Intuitively this could be measured by the values each agent assigns the next-states transitioned to by each one of the possible actions. The delta between the highest value action and the second highest can indicate how much better the agent thinks that action is, i.e. its confidence in its decision.

We'd like to argue that if an agent is confident in his choice of action, then this action should be his choice for similar states.

We dive into what constitutes state similarity in the following section, but for now let us assume we have a way of obtaining similar states given an origin state s_O s.t. $M(s_O, \epsilon) = \langle s_O^1, s_O^2, ..., s_O^n \rangle$, where ϵ is some parameter defining the reign of the state space we search.

3 State Similarity