Review-15

Evaluating Agents Without Rewards

Reinforcement Learning (RL) makes use of complex reward functions which may need to be hand-designed and engineered. Reward formulations are prone to human error and susceptible to sparsity. To this end, the work revisits objectives which enable the agent to learn without task rewards. Estimation and application of these objectives remains an open question and the work aims to throw light in this aspect. The study explores correlation between objectives and human behavior across seven agents and three Atari games.

A total of three RL objectives are considered which do not rely on task rewards. Firstly, curiosity encourages the agent to seek rare inputs using a learned density model. Empowerment, on the other hand, measures the anget's influence over its environment. Lastly, Information Gain rewards the agent as a rule discovery in the environment. The study compares objectives on Breakout, Seaquest and Montezuma's Revenge from the Atari benchmark by making use of PPO, ICM and RND agents. Metrics are additionally compared to pre-recorded human behavior, no-op and random agents. In order to pre-process pixel inputs, each a simple preprocessing scheme consisting of resizing and discretization is adopted to preserve the position of objects based on brightness percentiles. Images are then enumerated to represent each unique frame by an integer index. Indexing is carried out to evaluate discrete probabilities which are utilized in the agent's comparison with human behavior. Comparison between agents is carried dout using correlation metrics which assess the degree of similarity between objectives and task rewards, and objectives and human behavior.

Out of the three objectives considered, curiosity correlates strongly with human behavior. Task rewards with curiosity better explain task rewards alone. Empowerment and Information Gain correlate strongly with each other but demonstrate a weaker relationship to curiosity. While the study aims to understand tsk-agnostic behavior in RL, it lacks several essential components in its analysis. Firstly, the work only exploits probabilities based on state visitations during agent's lifetime which may not be a sufficient component. One could also employ attention mechanisms and kernel visualizations to monitor which pixels the agent finds important. In comparison to human behavior, curiosity is found to beneficial in presence of task rewards which highlights the necessity of task-rewards. The study does not provide insights into wide variations in results of empowerment and information gain and argues that these may vary as per the nature of the game. To this end, task-agnostic objectives may not necessarily scale to high-dimensional action spaces. Lastly, the simple preprocessing scheme and limited human data hinder the reliability of analysis in light of real-world RL applications.