Review-1

Action and Perception as Divergence Minimization

Divergence minimization has seen a tremendous growth in designing novel Reinforcment Learning (RL) objectives over the past few years. A good choice of RL objective is essential in maintaining a balance between informative exploration and inferring environment representations. To that end, the work presents a unified objective which combines the key elements of action and perception as joint divergence minimization. The key idea behind the objective is to extract relevant information from representations observed by the agent and utilize them in an efficient manner to align future dynamics as per the beliefs of the agent.

Various information-based objectives in literature such as inference methods, information gains, entropy-based exploration and free-energy minimization principle can be thought of as a joint divergence minimization approach. The unified objective consists of an actual distribution (which is a generative process parameterized by the agent) and a target distribution (which is desired by the agent). The joint divergence minimization makes use of the Kullback-Liebler (KL) divergence metric to bring the actual distribution closer to the target distribution by optimizing over the agent's parameters. The optimization of the agent consists of two stages. (1) Firstly, the perception phase aligns beliefs of the agent with its past inputs based on inference of latent variables. These latent variables are not observed over time and can only be utilized in an indirect manner such latent representations in representation learning or model parameters in the case of model learning. (2) Secondly, the action phased aligns the future inputs with the agent's beliefs by making use of the latent variables estimated in the perception. These typically are made use of in a future term entropy in exploration, skill discovery in hierarchical RL or information gain in model learning. The combinationg of action and perception renders task rewards optional since the agent itself learns in latent space.