Review-1

Action and Perception as Divergence Minimization

Divergence minimization has seen 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 gain, 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 phase aligns the future inputs with the agent's beliefs by making use of latent variables estimated in the perception phase. These typically are made use of in a future term such as 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.

The work presents a detailed review of various information-based objectives and their retrieval from the unified action-perception objective. For instance, the divergence minimization is a sound generalization of inference methods such as variational and amortized inference; exploration techniques such as maximum entropy RL and empowerment; and temporal abstraction such as skill discovery. Moreover, the review maintains a close connection between the actual and target distributions highlighted in the definition of the objective which provide a generalized procedure for constructing novel objectives for future work. This can be achieved by treating the target distribution as a probabilstic model under which the agent infers informative representations and explores future inputs. On the other hand, the unified objective itself is constructed of primitive entities observed in literature and does not give rise to a new objective. For instance, the objective is a theoretical collection of past objectives which are linked using the divergence minimization scheme. Moreover, the framework proposed is dense and throws little light on how to practically use the unified objective in the formulation of novel information-based methods. Lastly, the unified objective mentions connections to a few essential methods such as surprise minimization but does not take them into account as a significant constituent.

The work opens three new directions for future work. Firstly, utilization of latent variable models from a free-energy perspective provides insights into the theoretical usage and implementation of niche-seeking objectives. The question one may ask is can we utilize the free energy principle to design objectives for real world applications such as robotics? Secondly, theoretical unification of information schemes paves the way for practical applicability for constructing new objectives. Lastly, the divergence minimization itself can be broadened by adding novel elements to the actual distribution based on the past and future of the agent. For instance, with the reward being an optional entity in the agent's observation space, can it be effectively leveraged in latent space? If not, then what other entities of the agent's dynamics could be used for effective latent exploration? While the incorporation of new latent variables remains unclear, the unified objective succeeds in steering research towards this direction.