Paper Review: Variational Intrinsic Control

Summary:

The paper "Variational Intrinsic Control" introduces a method for learning control policies in an unsupervised manner. The method relies on optimizing an intrinsic reward signal that quantifies the extent of the agent's control over the environment. Also, the method is inspired by the concept of empowerment and is derived from the theory of optimal control. The authors demonstrate the effectiveness of their approach on a variety of tasks, including locomotion, manipulation, and navigation, within open and closed loop environments. It is shown that an agent learns to collect rewards significantly faster after given the opportunity to explore its environment first.

Contributions:

This papers main contribution is the framework for Variational Intrinsic Control (VIC) which combines variational autoencoders with intrinsic curiosity-driven exploration to enable agents to learn a low-dimensional latent representation of the environment and generate intrinsic rewards for exploration. Allowing the agent to explore its environment before being given a goal is a unique addition to a DRL model that has not been done before.

Strengths and Weaknesses:

The authors state, and I agree, that the major strengths of VIC is that it is relatively simple, uses closed loop policies, can use general function approximation, discrete and continuous, and is model-free. However, I believe that the potential weaknesses include its computational complexity, sensitivity to hyperparameters, limited generalization to diverse tasks, and potential overemphasis on intrinsic rewards. The authors suggest careful consideration and evaluation of these limitations when applying VIC in practical reinforcement learning scenarios.

Experimental Validity:

Two experiments were conducted in this paper, first with an agent that can move blocks around and second with the MINST number catalog. Open and closed loop options are tested, introducing or not introducing outside noise into the environment, and demonstrated that the convergence to the optimal policy remains the same. I believe that this experiment provides a strong and rigorous argument for unsupervised learning.

How can this work be extended:

The work in this paper could be extended to real-world applications, like the authors mentioned in the paper, if future work is done to make function approximation work by playing back previous experiences in the environment. Additionally, this work could be extended by implementing other strategies for exploration, such as incorporating external memory for more effective exploration strategies. Overall, it is very exciting to see what will come next with capabilities like the one described in this paper.