LEAF: Latent Exploration Along the Frontier

Real-world robot tasks require the use of efficient exploration strategies. These strategies are facilitated by goal proposal and temporally abstract schemes. However, these do not necessarily aid in discovering long horizon plans. To that end, the work proposes an exploration framework consisting of committed exploration along a localized set of states. Latent Exploration along the Frontier (LEAF) allows the robot to explore states near a frontier in latent space which has been previously explored. The LEAF policy consists of a deterministic component which places the robot at the frontier of states to be explored. The stochastic component, on the other hand, allows the agent to explore novel states. Simultaneous combinations of stochastic and deterministic executions result in committed exploration and efficient allocation of novel state budget. LEAF demonstrates improved exploration performance on challenging environments including a real robot task.

LEAF carries out exploration in latent space with pixel and state-based inputs. Corresponding to a frontier, the robot receives goal and input states. Reachability of the goal state from the initial state is evaluated using a $\beta - VAE$ which yields encodes both states and infers the frontier from the learned manifold. The VAE yields a binary valued output corresponding to whether the goal is reachable or not. In the case of reachable goals, the robot executes its deterministic policy to reach a state in the frontier and execute its stochastic policy to reach the final goal state. In order to facilitate reachability towards goals, the robot is made to sample easy-to-reach goal states towards the beginning of training using an implicit curriculum. Training additionally consists of $\beta - VAE$ which is optimized to minimize the distance between the initial and goal states.

LEAF carries out committed exploration as the set of states leading to the frontier have been previously explored. This results in an efficient exploration budget and allows the robot to exhibit goal-directed behavior while exploring simultaneously. Additionally, provision of an implicit goal-conditioned curriculum allows the robot to learn temporally abstract policies which aid in globally optimal behavior. On the other hand, the method can be improved in regard to its training and evaluation setup. LEAF assigns labels to VAE training data based on a threshold heuristic which may not necessarily generalize to real-world tasks. Moreover, the method does not tune the threshold parameter or provide any insights into its practical usage. A more suitable method for data labelling would be to learn the threshold using an auxilary architecture. Additionally, the method assumes that deterministic component of the robot policy is near-optimal and can place the robot precisely at the frontier location. The assumption comes out as too stong since the deterministic component is obtained from the exploration of stochastic component and may not be necessarily optimal with regard to frontier as a continuum of states.

LEAF lays out an exploratory curriculum which is composed of deterministic execution followed by stochastic learning. The mthod presents two new directions for future work. While the labelling of training data is based on an untuned heuristic, LEAF provides the opportunity to provide suitable labelling schemes for data on-the-fly. This would aid in effective training labels and accurate goal-directed behavior in latent space. Lastly, the method provides room for scalability to high-dimensional real robot tasks consisting of larger degrees of freedom and partial observability.