Invertible Q-Functions

Dereck Piche

August 8 2024

Modulation is best avoided in ML. There should not be a separation between the Q_θ function (the judge) and the action function U_ϕ (the executioner).

A simple way of avoiding this separation is to approximate Q by an easily invertible parameterized function Q_θ. The ease of inversion should remain after conditioning on any state s.

If Q_θ generalizes well the mapping Q of actions/states subspace (S, A) to a return subspace T, there should be a guaranteed generalization of Q_θ^{-1} from T to A if we condition on a $s \in S$.

Then $Q_\theta^{-1}(t; s)$ should yield a good action if $s \in S$, t is high and $t \in T$ and a in A. The key difficulty is to determine if the later two requirements are met.

If Q_θ parameterizes a distribution we can gradient descent to lower the entropy.

Another way is to have Q_θ map S, A to T, C where $c \in C$ measures confidence. The training algorithm would lower the divergence penalty if the confidence is low.

Then with high $h \in H$ and $c \in C$ the inversion will give a likelier good $a \in A$.