Modularity in Reinforcement Learning via Algorithmic Independence in Credit Assignment

ABSTRACT

Many transfer problems require re-using previously optimal decisions for solving new tasks, which suggests the need for learning algorithms that can modify the mechanisms for choosing certain actions independently of those for choosing others. However, there is currently no formalism nor theory for how to achieve this kind of modular credit assignment. To answer this question, we define modular credit assignment as a constraint on minimizing the algorithmic mutual information among feedback signals for different decisions. We introduce what we call the modularity criterion for testing whether a learning algorithm satisfies this constraint by performing causal analysis on the algorithm itself. We generalize the recently proposed societal decision-making framework as a more granular formalism than the Markov decision process to prove that for decision sequences that do not contain cycles, certain single-step temporal difference action-value methods meet this criterion while all policy-gradient methods do not. Empirical evidence suggests that such action-value methods are more sample efficient than policy-gradient methods on transfer problems that require only sparse changes to a sequence of previously optimal decisions.
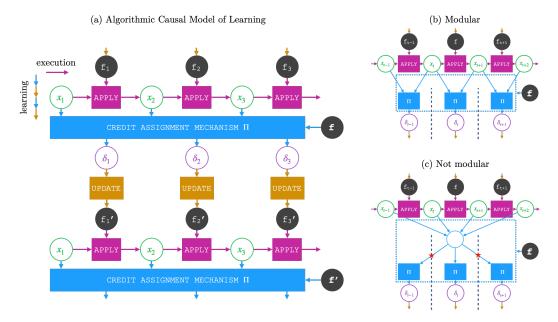
Figure 1: **Algorithmic Causal Model of Learning.** A learning algorithm with credit assignment mechanism Π that produces gradients δ to update functions $\mathtt{f}$ to $\mathtt{f}'$ can be represented as a causal graph (a). Π is not modular (b) if it contains a hidden variable whose outgoing causal edges cut a partitioning among the δ's (shown by the red star) and modular (c) if it does not.