Information theoretic approaches for predictive models: results and analysis

by Monica Dinculescu

Supervised by Prof. Doina Precup

Abstract

Learning the internal representation of partially observable environments has proven to be a difficult problem. State representations which rely on prior models, such as partially observable Markov decision processes (POMDPs) and hidden Markov models (HMMs), are computation expensive and sensitive to the accuracy of the underlying model dynamics. Recent work by Still and Bialek proposes an information theoretic approach that compresses the available history into an internal representation and maximizes its predictive power. This method also allows for the agent to act on the world and influence the observations received. A key challenge becomes maintaining a balance between taking advantage of the predictive power of the model while at the same time exploring unlikely future observations to improve the model. In order to validate the asymptotic nature of the theoretical algorithm, we present the first empirical results in the field, demonstrating the accuracy of the internal representation as well as its predictive powers. In addition, we propose two alternative approaches that may ensure faster convergence times and more accurate optimal action strategies.