Proc. Fourth Int. Conference on Cognitive and Neural Systems, Boston, 2000,p. 22

Benchmarking Reinforcement Learning based on Neural Function Approximators

D. Surmeli and H.-M. Gross
Dept. of Neuroinformatics, Ilmenau Technical University, Germany
(dima@informatik.tu-ilmenau.de)

Framework: Considering observable behavior as the only truly relevant criterion for an evaluation of algorithms for perception and generation of behavior, we are interested in sensorimotor problems of mobile robotics. With their characteristics (real world sensors and actuators, embodiment etc.), robots provide unique testbeds, but also equally valuable insights for biologically motivated approaches.
Motivation: We focus on such approaches to realize the sensorimotor paradigm: perception and action generation are to be fused into one generative process. An example of this unified process is temporal difference reinforcement learning (tRL) and especially, its Q-variants, since they utilize just one sensorimotor map, but also because they originally aimed at modeling conditioning as the base of natural intelligent behavior. The success in modeling classical conditioning, however, was not paralleled for operand conditioning. So, our final interest is in a biologically motivated model (bRL) of action generation.
Recently, research into the characteristics of tRL using neural function approximators intensified dramatically, yet no systematic comparison or methodology for such has been established. Moreover, a classification of the approaches is lacking. This contribution intends to suggest a few points towards these ends.
Intention: Present real world problems include continuous input and output spaces, instationarity and partial observability of the environment. However, the comparison of tRL solutions is executed in a simulated environment to accelerate investigations and reach conclusive results about factors relevant to performance. The inverted pendulum task was chosen as a first problem for a systematic investigation of the above problems and as a standard benchmark often used. Other suggested robotic tasks will be implemented to point out different aspects of the algorithms. Behavior as a real sensorimotor system remains the ultimate test.
For a classification and selection among the many algorithms, a few of the criteria to consider from a multidimensional structure include artificial (aNN) vs biologically motivated neural nets (bNN); learning on closed sets vs lifelong (where the stability/plasticity and with it in RL, the exploration/exploitation dilemma must be addressed); with aNN: unsupervised clusterers with an additional supervised layer vs direct function approximators; fixed vs incremental topologies; input driven vs task-relevant clustering and for those, statistical vs one shot learning for rare, important events.
Method: For a reinforcement learning task, each pattern in a set represents a starting point for a trial or trajectory until the agent reaches a goal or a maximum number of steps. The 3 mutually exclusive sets are comprised of randomly selected points, such that cross validation and multiple realizations become necessary. As in supervised benchmarks, the criterion to stop learning and enter the test set measures the development of adaptable parameters, e.g. Q-values. Then, performance may be measured by different suitable criteria.
Results: We report on the performance of algorithms involving a number of Q-learning variants (Q, Sarsa, Q(λ)) and input driven clustering (neural gas, growing neural gas) to discretize continuous inputs and featuring eligibility traces as a means to speed up learning, but also of dealing with partial observability. Alternatively, we investigate P- and C-Trace algorithms designed to cope with the latter problem.
Further, we examine the utility of exploiting the similarity computations of those clusterers to speed up learning among Q-values for similar state-actions pairs and find that it dramatically speeds up learning.
To enable one shot learning, those neural gas clusteres are extended to include an ART-like module that detects illegal dissimilarities and thus, captures outliers important for the sensorimotor mapping.
Conclusion: For sensorimotor tasks, we need lifelong adaptable algorithms capable of fast learning and capturing important singular, along with statistical, events. This lead us to neural approximators, especially clusterers from fixed (NG) to incremental topologies (GNG) to ART for fast statistical and singular learning. ARTMAPs are among the few networks with the desired characteristics, yet its variants remain to be applied with the changing target vectors typical for reinforcement learning.