A Modular Q-Learning Architecture for Manipulator Task Decomposition

Chen K. Tham*
Department of Engineering
University of Cambridge
Cambridge CB2 1PZ
United Kingdom
ckt@eng.cam.ac.uk

Richard W. Prager
Department of Engineering
University of Cambridge
Cambridge CB2 1PZ
United Kingdom
rwp@eng.cam.ac.uk

Abstract

Compositional Q-Learning (CQ-L) (Singh 1992) is a modular approach to learning to perform composite tasks made up of several elemental tasks by reinforcement learning. Skills acquired while performing elemental tasks are also applied to solve composite tasks. Individual skills compete for the right to act and only winning skills are included in the decomposition of the composite task. We extend the original CQ-L concept in two ways: (1) a more general reward function, and (2) the agent can have more than one actuator. We use the CQ-L architecture to acquire skills for performing composite tasks with a simulated two-linked manipulator having large state and action spaces. The manipulator is a non-linear dynamical system and we require its end-effector to be at specific positions in the workspace. Fast function approximation in each of the Q-modules is achieved through the use of an array of Cerebellar Model Articulation Controller (CMAC) (Albus 1975) structures.

NOTE

The full paper can be found in the main ML'94 proceedings.

RESEARCH WORK AND INTERESTS

Our research interests involve the scaling up of machine learning methods, especially reinforcement learning, for autonomous robot control. We are interested in function approximators suitable for reinforcement learning in problems with large state spaces, such as the Cerebellar Model Articulation Controller (CMAC) (Albus 1975) which permit fast, online learning and good local generalization.

In addition, we are interested in task decomposition by reinforcement learning and the use of hierarchical and modular function approximator architectures. We are examining the effectiveness of a modified Hierarchical Mixtures of Experts (HME) (Jordan & Jacobs 1993) approach for reinforcement learning since the original HME was developed mainly for supervised learning and batch learning tasks.

The incorporation of domain knowledge into reinforcement learning agents is an important way of extending their capabilities. Default policies can be specified, and domain knowledge can also be used to restrict the size of the state-action space, leading to faster learning. We are investigating the use of Q-Learning (Watkins 1989) in planning tasks, using a classifier system (Holland 1986) to encode the necessary condition-action rules.

References

Albus, J. (1975), 'Data storage in the cerebellar model articulation controller (CMAC)', *Journal of Dynamic Systems, Measurement and Control* **97**(3), 228–233.

Holland, J. (1986), Escaping brittleness: The possibilities of general-purpose learning algorithms applied to parallel rule-based systems, *in* R. Michalski, J. Carbonell & T. Mitchell, eds, 'Machine Learning: An Artificial Intelligence Approach', Vol. II, Morgan Kaufmann, Los Altos, CA, chapter 20, pp. 593–623.

Jordan, M. & Jacobs, R. (1993), Hierarchical mixtures of experts and the EM algorithm, Technical Report 9301, MIT Computational Cognitive Science.

Singh, S. (1992), 'Transfer of learning by composing solutions of elemental sequential tasks', *Machine Learning* **8**(3/4), 323–339.

Watkins, C. (1989), Learning from Delayed Rewards, PhD thesis, University of Cambridge, Cambridge, UK.

*From January 1995: Department of Electrical Engineering, National University of Singapore. e-mail: eletck@leonis.nus.sg