Reinforcement Learning of POMDPs

Using a Discrete Hierarchical Self-Organizing Map

Georgios Pierris and Torbjorn S. Dahl

Cognitive Robotics Research Centre, University of Wales, Newport

georgios.pierris | torbjorn.dahl @newport.ac.uk

University of Wales, Newport

Prifysgol Cymru, Casnewydd

Introduction

Reinforcement Learning Self-Organizing Map (RLSOM) is a novel reinforcement learning approach based on a hierarchical structure, whose building blocks are augmented Kohonen Self-Organizing Maps.

Motivation

RL algorithms have shown remarkable results in static, discrete environments, e.g., agents playing chess. However, complex problems, e.g., humanoid robot control, lie in high dimensional continuous state-action environments.

RLSOM

✓Reduces the input space ✓Encodes it in fixed sequences

✓Increasingly abstract representations towards the top

✓Learns complex trajectories with perceptual aliasing

The bottom layer serves as state and action space reduction algorithm, whereas the higher layers encode activation patterns producing increasingly abstract representations towards the top of the hierarchy. Activating winning nodes considers the input match (IM), the match between short-term and long-term memory (HM), and the discounted future reward (DR). The activation potential of a node is the sum weighted according to our needs:

$$AP = w_1 HM + w_2 IM + w_3 DR$$

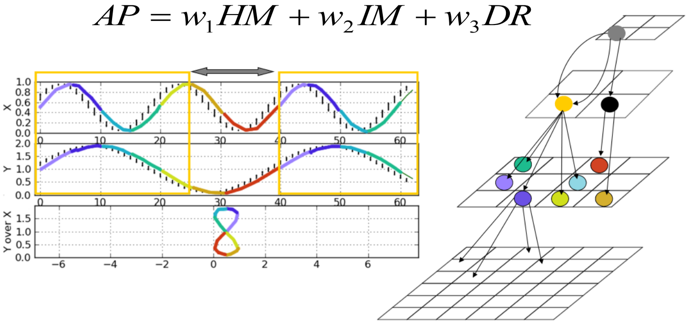

Figure 1: Each coloured sub-pattern of the training data (left) is encoded in the corresponding nodes (right). Sequences of sub-patterns are encoded higher and allows for reuse. Note the yellow node at level 2 reactivated from the grey top node.

Experiments and Results

In a rather complex experiment, a part of the Cornu Spiral is used to artificially produce a loop, i.e., a hidden state.

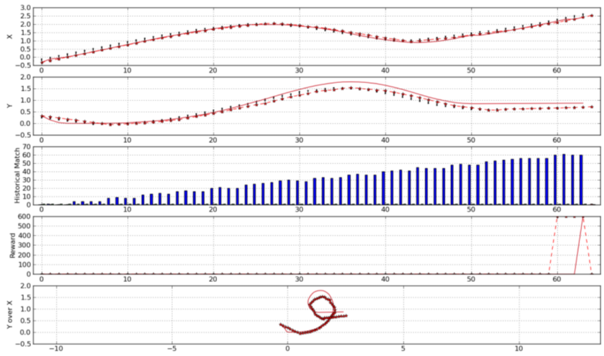

Figure 2: The red line is the original data and the stars represent the reproduction. X, and Y, are time-series, and the bottom figure shows Y over X to highlight the hidden state. The blue bars indicate the increasing historical match over time.

RLSOM chooses between better fit of past experiences (Figure 2) or balances history and reward (Figure 3) to use nodes in different context, e.g., follow shorter path to goal.

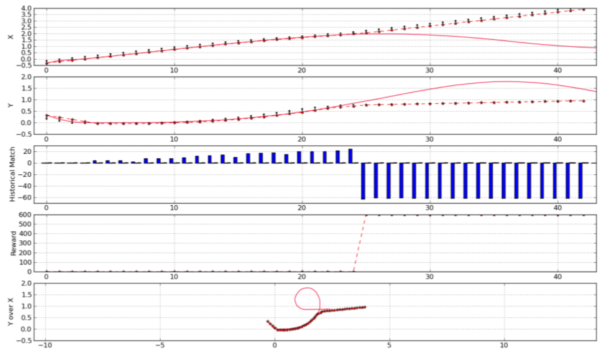

Figure 3: Note the rapid change in the policy at step 25. The Historical Match drops and RLSOM follows the shorter path to the reward.

Conclusions

✓RLSOM discretizes the input space and successfully encodes temporal sequences in a hierarchical SOM.

✓Problems that include hidden states can be learned and reproduced under different contexts.

Acknowledgments:

The research leading to these results has received funding from the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement ROBOSKIN ICT-FP7-231500.

RoboSkin

References:

L. B. Cohen, H. H. Chaput, and C. H. Cashon, ``A constructivist model of infant cognition'', *Cognitive Development*, vol. 17, no. 3-4, 2002

A. Shimada and R. Taniguchi, ``Gesture recognition using sparse code of hierarchical SOM'' in *ICPR*, 2008

CRRC
COGNITIVE ROBOTICS RESEARCH CENTRE