Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 14 (2012) 10–25

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Motivated learning for the development of autonomous systems

Action editors: Minho Lee and Wlodzislaw Duch

Janusz A. Starzyk [a], James T. Graham [a,*], Pawel Raif [b], Ah-Hwee Tan [c]

[a] School of Electrical Engineering and Computer Science at Ohio University, Athens, OH, USA
[b] Faculty of Organization and Management, Silesian University of Technology, Gliwice, Poland
[c] School of Computer Engineering, Nanyang Technological University, Singapore

Received 1 April 2010; accepted 12 December 2010
Available online 12 January 2011

Abstract

A new machine learning approach known as motivated learning (ML) is presented in this work. Motivated learning drives a machine to develop abstract motivations and choose its own goals. ML also provides a self-organizing system that controls a machine's behavior based on competition between dynamically-changing pain signals. This provides an interplay of externally driven and internally generated control signals. It is demonstrated that ML not only yields a more sophisticated learning mechanism and system of values than reinforcement learning (RL), but is also more efficient in learning complex relations and delivers better performance than RL in dynamically-changing environments. In addition, this paper shows the basic neural network structures used to create abstract motivations, higher level goals, and subgoals. Finally, simulation results show comparisons between ML and RL in environments of gradually increasing sophistication and levels of difficulty.

Keywords: Autonomous systems; Intelligent agents; Motivated learning; Neural networks; Reinforcement learning

1. Introduction

Intelligent machines are expected to revolutionize the way we live, yet we still do not know how to design or build systems with "true" intelligence. The biological brain is both an inspiration and a model for the development of intelligent machines. We cannot build a brain, but we can try to design models that exhibit similar activation of perceptions, memories and motor control in a given environment. Artificial neural networks (ANN) inspired by networks of biological neurons are successfully used for classification, function approximation and control. Yet a classical ANN learns only a single task, requires extensive training effort, and close supervision.

The reinforcement learning (RL) mechanism is related to the way animals and humans learn (Bakker & Schmidhuber, 2004). Based only on occasional reward and punishment signals, RL agents must learn how to interact with their environment to maximize their expected reward. However, the learning effort and computational cost increase significantly with the environmental complexity (Barto & Mahadevan, 2003), thus, optimal decision making in a complex environment is still intractable using RL. This feature, usually called "the curse of dimensionality", is one of the main disadvantages of RL in real-world applications.

Reinforcement learning also suffers from what is called the "credit assignment problem" (Sutton, 1984; Fu & Anderson, 2006). Reinforcement learning uses a temporal difference mechanism to spread the value of the reward received to earlier stages. However, it does not have a natural mechanism to stop the spread of the reward to yet earlier stages that had nothing to do with receiving the reward. O'Reilly proposed a new primary value and

* Corresponding author.
E-mail addresses: starzykj@gmail.com (J.A. Starzyk), jg193404@ohio.edu (J.T. Graham), pawel.raif@polsl.pl (P. Raif), asahtan@ntu.edu.sg (A.-H. Tan).

1389-0417/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2010.12.009

learned value (PVLV) scheme that implements Pavlovian conditioning (O'Reilly, Hazy, Watz, & Frank, 2007). PVLV directly associates the stimuli and the reward and is a promising alternative to the temporal-differences (TD) used in traditional RL (O'Reilly & Frank, 2006).

One way to improve the efficiency of RL is to use subgoals to build a hierarchy of subsequent goals. The hierarchical reinforcement learning (HRL) approach tends to exploit the structure of both the environment and the agent's tasks to improve policy learning in large scale problems. Among the many approaches to hierarchical RL one can distinguish: Dayan and Hinton's research on feudal reinforcement learning (Dayan & Hinton, 1993), the study by Parr and Russell (1998) on hierarchical abstract machines (HAM) and development of MAXQ Method (Dietterich, 2000).

Bakker and Schmidhuber (2004) proposed a method for hierarchical reinforcement learning based on subgoal discovery and subpolicy specialization. Their HASSLE algorithm can outperform plain RL "by learning to create both useful subgoals and the corresponding specialized subtask solvers." In their algorithm they use HASSLE (Harmon & Baird, 1996) on both high and low levels of hierarchy. Among the limitations of this system are the large number of parameters, the lack of strict convergence guarantees and the dependence on identifying reasonable high-level observations.

Subgoals discovered in hierarchical reinforcement learning (HRL), are obtained by clustering input data (Bakker & Schmidhuber, 2004) to arrive at desired and useful results. In HRL, high-level policies are used to discover subgoals and apply them when appropriate to accomplish the goal. This yields automatic learning of the goal hierarchy minimizing the designer's effort. High-level policies optimize the subgoals and manage their real time use. Individual subgoals are managed by low-level policies that learn low-level value functions in the sensory–motor subspaces. However, identification of useful subgoals is not easy and the large number of design parameters limits the usefulness of the HRL method. While HRL with subgoal discovery does improve machine learning, it still suffers from the major limits of RL, since it is focused on maximizing total reward for externally set objectives.

However, what if we ascribe motivations to machines? An intelligent machine must be able to generate and pursue goals on its own, learning what it needs for a given set of assigned tasks, exploring for a reason, developing new motivations and setting its own goals. Existing methods have made some progress in this direction (Bakker & Schmidhuber, 2004; Barto, 2004; Huang & Weng, 2002; Oudeyer, Kaplan, & Hafner, 2007, 2010; Roa, Kruijff, & Jacobsson, 2009; Schmidhuber, 1991)

The key question is how to "motivate" a machine to act and enhance its intellectual abilities, how to improve its learning efficiency, and how to design a mechanism for structural self-organization from which higher level perceptions and skills could evolve through the machine's interaction with its environment (Pfeifer & Bongard, 2006; Steels, 2004)? What can drive an agent to explore the environment and learn the ways to effectively interact with it? Finally, how can a machine be designed that is capable of not only implementing given goals but also creating them and deciding which goals to pursue? How can this be done in a constantly changing environment, and in spite of distractions and unforeseen difficulties?

1.1. Intrinsic motivation and curiosity driven exploration

According to Pfeifer and Bongard (2006), an agent's motivation should emerge from the developmental process. This is observed in humans and has been argued that it is the result of a system that rewards the engagement of activities just above a person's current ability level. Humans seem to have an innate need to ask "Why?" and "How?" in order to understand the world.

Based on the curiosity principle, Oudeyer et al. (2007, 2010) proposed an intelligent adaptive curiosity (IAC) system, which attempted to direct a robot in continuous, noisy, inhomogeneous, environments, allowing for an autonomous self-organization of behavior toward increasingly complex behavioral patterns. It is widely believed that intrinsic motivation is integral to the way humans learn and explore their environment (Cohn, Ghahramani, & Jordan, 1996; Hasenjäger & Ritter, 2002; Schmidhuber, 1991; Schultz, 2002; White & September, 1959). Oudeyer discusses the benefits children gain by exploring their environment and some of the reasoning behind such behavior (Oudeyer et al., 2007, 2007). Development in children is considered to be autonomous and active, and while adults can provide assistance, it is only assistance. The children's decisions are (largely) their own. The fact that children like to play, and that they actively choose to play for the sake of play, rather than as a step toward solving practical problems, can be taken as proof of the existence of a kind of intrinsic motivation system.

Roa et al. (2009) explored the concept of curiosity and whether it can be emulated through a combination of active learning and RL using intrinsic and extrinsic rewards. The authors developed their intrinsic motivation system based on Oudeyer's work (Oudeyer et al., 2007), and then added an extrinsic reward system to guide the robot to its goal.

By using a learning mechanism based on intrinsic motivations, a machine can explore the environment and learn a hierarchy of skills that it will need to work in this environment (Barto, 2004). Intrinsic motivation can be based on surprise, novelty (Huang & Weng, 2002), or a learning progression as discussed by Kaplan and Oudeyer (2004).

Intrinsic motivation as used in curiosity based learning is similar to exploration in reinforcement learning. In RL a machine does not always respond in an optimum way but occasionally tries a random search in state-action space. However, without proper oversight of curiosity