Information Sciences 180 (2010) 2856–2874

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Empirical analysis of an on-line adaptive system using a mixture of Bayesian networks

Daisuke Kitakoshi [a,*], Hiroyuki Shioya [b], Ryohei Nakano [a,1]

[a] Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
[b] Computer Science and Systems Engineering, Muroran Institute of Technology, Mizumoto-cho, Muroran 050-8585, Japan

ARTICLE INFO

Article history:
Received 17 April 2008
Received in revised form 20 January 2010
Accepted 2 April 2010

Keywords:
Adaptation to dynamic environments
Mixture of Bayesian networks
Reinforcement learning
Profit sharing

ABSTRACT

An on-line reinforcement learning system that adapts to environmental changes using a mixture of Bayesian networks is described. Building intelligent systems able to adapt to dynamic environments is important for deploying real-world applications. Machine learning approaches, such as those using reinforcement learning methods and stochastic models, have been used to acquire behavior appropriate to environments characterized by uncertainty. However, efficient hybrid architectures based on these approaches have not yet been developed. The results of several experiments demonstrated that an agent using the proposed system can flexibly adapt to various kinds of environmental changes.

1. Introduction

Intelligent systems through which robots and agents can effectively learn new behaviors have been actively researched and developed in various fields. One recent research effort in particular focused on building an intelligent system that can adapt to uncertain and dynamically changing environments [2,29]. Reinforcement learning (RL) is a kind of machine learning that can be applied in such environments [50]. Many RL methods have helped improve learning performance in environments modeled by Markov decision processes (MDPs), and some of these methods have also been applied to problems in non-MDP domains [8,11,16,32,33,35–37,46,47,51,52,58,59]. A typical example of a non-MDP model is the partially observable MDP (POMDP) model [24]. In environments modeled as POMDPs, learners do not have complete information regarding their current state due to the presence of noisy input and the detection limit of the learner's sensor. Stochastic systems can be used to acquire reasonable behavior in environments characterized by uncertainty. A Bayesian network (BN), a kind of stochastic model, can provide appropriate noise-robust output through probabilistic inference [14,20,21,31,39,40,42,48,49]. BNs are capable of representing models in various problem domains and are applicable to a broad range of problems: automatic driving control [12], behavior control for robots [22,62], and so on [6,9,26,44]. In real-world environments, both the display of appropriate behavior with respect to unobservable input and adaptability to changes in the environment are required.

In this paper, we consider situations in which the environment is discretely changed at fixed time intervals. In the field of RL, several improvements have been devised concerning such situations, such as (i) modifying and reusing policies related to

* Corresponding author. Present address: Department of Computer Science, Tokyo National College of Technology, 1220-2 Kunugida-machi, Hachioji, Tokyo 193-0997, Japan. Tel./fax: +81 42 668 5061.
E-mail addresses: kitakosi@tokyo-ct.ac.jp (D. Kitakoshi), shioya@csse.muroran-it.ac.jp (H. Shioya), nakano@cs.chubu.ac.jp (R. Nakano).
[1] Present address: Department of Computer Science, Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi 487-0027, Japan.

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.04.001

previous environments [35,55,56] and (ii) taking advantage of a single policy chosen from a policy library [11]. Nevertheless, there has been little research thus far on applying stochastic model-based systems to adaptation to dynamic environments.

Let us consider how people adapt to changes in the environment. People generally store "knowledge", i.e., experience, about a number of already solved problems and can solve a problem with high probability if the same problem has previously been encountered and solved. In addition, people can generally solve problems slightly different from previously encountered ones by utilizing their experiences to suitably modify previous solutions. Even completely new problems can be treated more effectively by mixing reusable experiences concerning previously encountered problems. Incorporating this ability into adaptive systems so that they can adapt to a variety of environments is the objective of this study.

Inspired by the human ability to act on the basis of empirical knowledge, we focused on how to treat and leverage these experiences for application to agents' (robots') behavior learning, and we developed an on-line system capable of adapting to environmental changes by using a mixture of BNs [28]. In the proposed system, called the *IPMBN* (*on-line adaptive system for Improving agents' Policies using a Mixture of BNs*), the learning object is regarded as an RL agent, while the object's knowledge regarding a certain behavior in an environment, called a policy, is represented by a mixture distribution of BNs. The empirical data from the RL agent is presented as sequences of the agent's states, actions, and rewards. The BNs store the data statistically, and the experiences are presented as a probability distribution. Each BN in the mixture, therefore, represents the stochastic characteristics of an individual agent's policy in the respective environment. The mixtures can provide policy information not only about previously encountered environments or similar ones but also about environments without any corresponding BNs in the mixture (called "unencountered environments").

From the theoretical viewpoint, a distinctive mixture formulation (called the "exponential mixture") is introduced to the IPMBN. A standard mixture formulation (called the linear mixture) is also introduced for comparison. In the proposed system, the assumptions the agent makes about its current environment are represented by a mixture of BNs. When environmental changes are observed, the system modifies the mixture to adapt to the changed environment. It then improves the agent's policy using the information represented in the mixture. Knowledge about the agent's behavior in the environment may be modified and partially utilized or may occasionally be used as a negative example, depending on the current environment. This corresponds to how people adapt to changes in the environment.

In real-world applications based on the mixing of reusable experiences, it is difficult to endow agents (or robots) with an ability to adapt to environmental changes. Therefore, the use of such mixing in real-world experiments has not been conclusively demonstrated. Actual applications using the proposed system in the real-world (for example, learning effective behavior in robot soccer, or adaptive and intelligent control of objects such as automobiles) have also not yet been realized. Nevertheless, this paper presents experiments using a mobile robot and computer simulations as trial cases. The results show that policy-improvement using a mixture of BNs enables an RL agent to adapt more flexibly to various kinds of environmental changes by modifying, reusing, and combining its knowledge. Moreover, the IPMBN displays acceptable performance even in the following problematic cases:

(i) where it is difficult to reuse the policies learned for previous environments for the current one;
(ii) where the current environment differs significantly from the ones the agent previously encountered.

The paper is organized as follows. Section 2 briefly explains an RL method known as profit sharing as well as the BN concept. Section 3 describes the process of representing the agent's policy information using BNs in the IPMBN, the concept of a BN mixture, and the algorithm governing the adaptation of the IPMBN to dynamic environments. Section 4 describes computer simulations conducted to evaluate the system's characteristics and performance, presents the results, and discusses their implications. It also describes experiments using a mobile robot, presents those results, and discusses them. Section 5 discusses related work. Finally, the key points are summarized and future research is mentioned in Section 6.

2. Preliminaries

This section reviews the two principal components of the proposed system: the reinforcement learning (RL) method, called "profit sharing", and the Bayesian networks (BNs). Profit sharing enables agents to learn new behaviors, while BNs represent knowledge regarding the agent's behavior in the environment.

2.1. Profit sharing

An RL agent[2] attempts to acquire an appropriate policy (a state-to-action map) on the basis of observations and trial-and-error interactions with its environment. A policy characterizes an agent's behavior. Various RL methods have been proposed and revised. Some of these, such as Q-learning [41,43,59,61], aim to optimize the total discounted reward (*value function*) using the concept of dynamic programming, typically in MDP environments. Other types of RL methods are based on the notion of credit assignment and aim to increase learning speed and to acquire reasonable policies. Profit sharing [16], a method of the latter type, is used as the policy learning mechanism in the IPMBN.

[2] In this paper, "agent" means any type of learning object, such as a mobile robot, vehicle, or any kind of forecasting system.