Research article

A robust cognitive architecture for learning from surprises

Thomas Joseph Collins *, Wei-Min Shen

Information Sciences Institute, University of Southern California, Los Angeles, CA, USA

A R T I C L E   I N F O

Article history:
Received 26 May 2017
Accepted 6 July 2017
Available online xxxx

Keywords:
Active learning
Prediction
Surprise-based learning

A B S T R A C T

Learning from surprises is a cornerstone for building bio-inspired cognitive architectures that can autonomously learn from interactions with their environments. However, distinguishing true surprises – from which useful information can be extracted to improve an agent’s world model – from environmental noise is a fundamental challenge. This paper proposes a new and robust approach for actively learning a predictive model of discrete, stochastic, partially-observable environments based on a concept called the Stochastic Distinguishing Experiment (SDE). SDEs are conditional probability distributions over the next observation given a variable-length sequence of ordered actions and expected observations up to the present that partition the space of possible agent histories, thus forming an approximate predictive representation of state. We derive this SDE-based learning algorithm and present theoretical proofs of its convergence and computational complexity. Theoretical and experimental results in small environments with important theoretical properties demonstrate the algorithm’s ability to build an accurate predictive model from one continuous interaction with its environment without requiring any prior knowledge of the underlying state space, the number of SDEs to use, or even a bound on SDE length.

© 2017 Elsevier B.V. All rights reserved.

Introduction

The ability to learn from surprises, which are mismatches between expected and actual sensory feedback in response to actions, is a cornerstone for building bio-inspired cognitive architectures that can learn autonomously from interactions with unknown environments. Theories and experiments in child psychology (Koslowski & Bruner, 1972; Piaget, 1953) and even the philosophy of science (Peirce & Hetzel, 1878) have emphasized the importance of surprises in the learning process. Experiments with AI agents that learn from surprises in biology (gene discovery, Shen, 1989; Shen & Simon, 1993), game playing (Shen, 1993a), learning from large knowledge bases (Shen, 1992b), learning on real robots in response to unexpected changes (Ranasinghe & Shen, 2008), and learning to detect and recover from interference and gaps in sensory information (Ranasinghe, 2012) have demonstrated the power of such a biologically-inspired approach. However, these existing techniques use logic-based rules to predict the results of future observations and modify their models in response to each prediction failure, which is a single mismatch between expected and actual observation. This leads to severe overfitting in noisy environments. Additionally, hidden-state detection and modeling (Shen, 1993b) relies on properties that are only satisfied in deterministic environments.

In this paper, we adopt a probabilistic approach that broadens the definition of a surprise to changes in the frequencies of individual prediction failures, leading to a novel and robust approach for actively learning a predictive model of discrete, stochastic, partially-observable environments (represented as Partially-observable Markov Decision Processes (POMDPs)) based on a novel concept called Stochastic Distinguishing Experiments (SDEs), which are probability distributions over the next observation given a variable-length ordered sequence of actions and expected observations up to the present. SDEs are simultaneously used to partition the space of possible histories the learning agent could encounter (thus forming an approximate predictive representation of state) and discover structure in sequences of actions and observations that allow the agent to predict future observations accurately. We derive this learning algorithm and present theoretical proofs of its convergence and computational complexity. Theoretical and experimental results demonstrate the algorithm’s ability to build an accurate predictive model from one continuous interaction with its environment without requiring prior knowledge of the underlying state space, the number of SDEs needed, or even a bound on the number of actions and observations in any SDE.
Related work

Much learning work regarding discrete POMDPs is in the area of reinforcement learning (RL), where the goal is not to construct a state representation but rather to learn an optimal policy given a known state space. Traditional exact and approximate techniques include Murphy (2000), Pineau, Gordon, and Thrun (2003) and Roy, Gordon, and Thrun (2005). SDEs share similarities with instance-based (IB) RL methods (Liu, Jin, & She, 2016; McCallum, 1995; McCallum, 1996; McCallum, Tesauro, Touretzky, & Leen, 1995), which memorize interactions with their environment and organize them into a suffix tree of actions and observations that approximates an unknown state space nonparametrically. SDE learning requires no memorization of instances (reducing required memory and avoiding the difficult issue of determining sufficient memory size), and, in contrast to notable IB methods (e.g., Zheng, Cho, & Quek, 2010), does not require separate training episodes or an externally-generated reward function. There exist techniques for growing a latent (unobserved) state space for unknown POMDPs based on an agent’s interaction with its environment that do not require specifying the number of states as prior information, most notably the infinite Partially-observable Markov Decision Process (ipOMDP, Doshi-Velez, 2009). Like other Bayesian approaches, specifying priors in ipOMDPs may be difficult in unknown environments, and there are strong modeling assumptions about these priors. One crucial difference between ipOMDPs and SDEs is that SDEs do not require an external reward function for action selection and learning. Additionally, as Littman notes in Littman, Sutton, and Singh (2002), predictive models (such as SDEs) built in terms of an agent’s actions and quantities it can directly observe have the potential to be easier to learn and generalize better.

Predictive state representations (PSRs, Littman et al., 2002) model the state of POMDPs in terms of the probabilities of tests (experiments), which consist of ordered actions and expected observations. For every controlled dynamical system, there exists a linearly independent set of tests to represent its state (i.e., is a sufficient statistic). Algorithms exist for finding a set of tests that forms a PSR (McCracken & Bowling, 2006; Wolfe, James, & Singh, 2005) and for learning the parameters of a set of tests forming a PSR (McCracken & Bowling, 2006; Singh, Littman, Jong, Parode, & Stone, 2003; Wolfe et al., 2005), which generally must be performed separately. Traditional algorithms for finding a set of tests forming a PSR generally proceed in a passive guess-and-check fashion, where subsets of potential tests are examined for linear independence. More recent techniques (Boots, Siddiqi, & Gordon, 2011; Hamilton, Fard, & Pineau, 2014) essentially circumvent the issue of learning the set of tests by maintaining a large (likely redundant) set of tests or resort to an offline, stochastic search (Liu, Zhu, Zeng, & Dai, 2016). In contrast, SDEs, which also build a model from the probabilities of experiments an agent can perform, can be selected and their parameters learned in an integrated fashion. Additionally, SDEs need not be linearly independent because they partition possible histories, saving an expensive linear independence check at each step. This also makes it easier to incrementally grow larger SDEs from smaller ones or merge SDEs together as the agent’s experience indicates a more (or less) complex representation is needed.

Finally, we note that SDEs are related to some techniques in the compression literature, such as Variable Order Markov Models (VMMs) for prediction (Begleiter, El-Yaniv, & Yona, 2004; Dimitrakakis, 2010a) and recent extensions of these methods to controlled processes (Dimitrakakis, 2010b) for online Bayesian inference and decision making. Traditional VMM approaches do not consider actions and the extensions to controlled processes select actions based on an externally-generated reward function, rather than actively generating actions to try to improve the agent’s model of the world (as in SDE learning). Additionally, the priors used in Bayesian approaches such as Dimitrakakis (2010b) may be difficult to set or even detrimental when the environment is completely unknown, as is the case in this work.

In summary, SDE learning has several key advantages in learning discrete POMDPs: (1) SDE learning is active, with the agent selecting and performing experiments in its environment to look for surprises that can be analyzed to provide crucial information about improving its model of the world; (2) SDE learning does not require specifying prior distributions or knowledge about the underlying state space, the number of SDEs, or a bound on SDE length; (3) SDE learning requires no external reward function because increases in estimated predictive model accuracy are used as a self-generated reward/goal; (4) SDEs can be learned via a single sequence of interactions with the environment (with no separate training episodes required).

**Stochastic Distinguishing Experiments (SDEs)**

**Stochastic Distinguishing Experiments (SDEs)** are an extension of Shen’s Local Distinguishing Experiments (LDEs, Shen, 1993b) for deterministic, partially-observable environments to stochastic environments. LDEs are ordered sequences of actions and expected observations that disambiguate states sharing the same observation. As an example, consider the shape environment first presented in Rivest and Schapire (1993) and shown in Fig. 1, in which an autonomous agent can perform actions \( x \) and \( y \) and receive observations \( square \) (in states I and II) and \( diamond \) (in states III and IV). The agent has no knowledge of the underlying state space or the consequences of its actions. In the deterministic version of this environment, states I and II can be disambiguated by observing \( square \), taking action \( y \), observing \( diamond \), and, finally, taking action \( x \), which, from state I, leads to a final observation of \( square \) and, from state II, leads to a final observation of \( diamond \). Thus, the LDE \( (square, y, diamond, x) \), informs the agent which of the two states with observation \( square \) it was at before it performed the LDE. The CDL algorithm (Shen, 1992a) incrementally builds a predictive model consisting of the LDEs and transitions between them necessary to distinguish all hidden states. These LDEs form a predictive representation of the state space.

When learning a set of LDEs, CDL assumes that executing the same trajectory from the same state any number of times will result in the same final observation and uses the historical differences in trajectories originating in different states to create and

---

**Fig. 1.** The shape environment, reproduced with kind permission from Shen (1994).
دریافت فوری
متن کامل مقاله

| امکان دانلود نسخه تمام متن مقالات انگلیسی | ✓ |
| امکان دانلود نسخه ترجمه شده مقالات | ✓ |
| پذیرش سفارش ترجمه تخصصی | ✓ |
| امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله | ✓ |
| امکان دانلود رایگان ۲ صفحه اول هر مقاله | ✓ |
| امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب | ✓ |
| دانلود فوری مقاله پس از پرداخت آنلاین | ✓ |
| پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات | ✓ |