Robotics and Autonomous Systems 56 (2008) 879–890

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

The initial development of object knowledge by a learning robot

Joseph Modayil [a,*], Benjamin Kuipers [b]

[a] Department of Computer Science, University of Rochester, PO Box 270226, Rochester, NY 14627, USA

[b] Department of Computer Sciences, The University of Texas at Austin, 1 University Station, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Available online 19 August 2008

Keywords:
Developmental robotics
Object representation
Object perception
Object actions
Machine learning

ABSTRACT

We describe how a robot can develop knowledge of the objects in its environment directly from unsupervised sensorimotor experience. The object knowledge consists of multiple integrated representations: trackers that form spatio-temporal clusters of sensory experience, percepts that represent properties for the tracked objects, classes that support efficient generalization from past experience, and actions that reliably change object percepts. We evaluate how well this intrinsically acquired object knowledge can be used to solve externally specified tasks, including object recognition and achieving goals that require both planning and continuous control.

1. Introduction

One reason why people function well in a changing environment is their ability to learn from experience. Moreover, learning from sensorimotor experience produces knowledge with semantics that are grounded in this experience. Replicating this human capability in robots is one of the goals of the robotics community. This paper describes how a robot can acquire knowledge of objects directly from its experience in the world.

This experiential knowledge has some significant advantages over knowledge directly encoded by programmers as it intrinsically captures the capabilities and limitations of a robot platform. The advantage of experiential knowledge has been demonstrated in the field of robot mapping where robot generated maps are more effective than human generated maps for robot navigation [31]. A similar situation arises when reasoning about objects, namely the robot's perception of the environment can differ greatly from that of a person. Hence, a robot should autonomously develop models for the objects in its environment, and then use these models to perform human specified tasks.

An important capability for a robot is to solve a current problem with knowledge acquired from past experience. Much research effort is spent on generalizing from past experience across *individual objects* ("a chair"), however a more pressing problem for a robot is to generalize across *experiences of individual objects* ("the red lab chair") so as to reliably reason and interact with the individual objects encountered in the environment over extended periods of time. This approach has the advantage that broader object classes can potentially be formed by weakening the restrictions on recognizing individual objects.

We describe how a physical robot can learn about objects from its own autonomous experience in the continuous world. The robot develops an integrated system for tracking, perceiving, recognizing, and acting on objects. This is a key step in the larger agenda of developmental robotics, which aims to show how a robot can start with the "blooming, buzzing confusion" of low-level sensorimotor interaction, and can learn higher-level symbolic structures of common-sense knowledge. We assume here that the robot has already learned the basic structure of its sensorimotor system [24] and the ability to construct and use local maps of the static environment [31].

The robot represents its knowledge of the individual objects in its environment with *trackers, percepts, classes* and *actions*. *Trackers* separate the spatio-temporal sensory experience of an individual object from the background. This sensory experience is filtered though perceptual functions to generate informative *percepts* such as the distance to the object and the object's shape. The robot uses the observed shape of a tracked object to generate shape *classes*, which the robot uses to efficiently generalize from past experiences. Finally, the robot is able to interact with an individual object using learned *actions* that modify the object's percepts.

In the following sections, we describe the motivations for this work, the representations for the objects and actions, the algorithm for learning this knowledge from experience, the evaluation with a physical robot, future directions and related work.

2. Motivation

Objects play a central role in the way that people reason about the world, so it is natural to want robots to share this capability.

* Corresponding author.
E-mail addresses: modayil@cs.rochester.edu (J. Modayil), kuipers@cs.utexas.edu (B. Kuipers).

0921-8890/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2008.08.004

However, the manner in which people think of objects is often different from the needs of a robot. People rarely have difficulty with object tracking, recognition, or interaction. However, it is difficult for a robot to acquire these capabilities even individually and a robot must be able to integrate these capabilities to solve tasks.

The semantic dictionary Wordnet [6] treats objects primarily as physical entities that belong somewhere on a classification tree. Wordnet attempts to bridge the gap between the representation of a word as a sequence of characters and its human-defined meaning by forming hierarchical relationships between words. For example, Wordnet states that a can is a container, while spoons and forks are cutlery. The Open Mind Indoor Common Sense project [9] goes further by defining multiple relationships between the names for common household entities. This style of knowledge is of little direct use to a robot without a connection between these words and the robot's experience in the world.

There is a broader role for object semantics, which is to support the formation of object representations that are defined from the robot's sensorimotor experience. It is tempting to think of objects as physical entities in the world that are derived from abstract classes, i.e. that the experience of seeing a fork is coming from a physical fork, which in turn is an instantiation of an abstract fork model that is shared by all people. The reality is the reverse, people start from sensorimotor experience, and classes are formed from individual experiences.

Instead of considering an object to be a physical entity, we consider an object to be an explanation for some subset of an agent's experience. With this approach, the semantics of an object are intrinsically defined from the agent's sensorimotor experience. When the robot uses its internal representations to solve externally specified tasks, the internal object representations may acquire a societally shared meaning.

Our approach to learning object models is inspired by theories in child development, in particular the assumption that coherent motion is one of the primary mechanisms for the initial perception of objects [28]. As modern techniques in robotics can effectively model the local static structure of the environment, any dynamic changes in the environment must come from some dynamic entity. By relying on a static environmental model, the robot can still perceive objects that are not moving. We use this as a basis for focusing our attention on learning about dynamic objects. The focus on dynamic objects provides a tractable way to make progress on an otherwise difficult problem—to perceive objects that have never been previously observed.

The focus of this work is to demonstrate how a robot can acquire an integrated set of object representations that can be used to solve externally specified tasks although the representations are internally formed directly from the robot's experience in the world. If a human and a robot are to share similar meanings for objects, then the robot must be able to perceive previously unseen physical objects, reason about the perceived objects, and take actions to achieve goals. The efficacy of the robot's internal object representations can be measured by how well they enable the robot to accomplish externally specified tasks such as recognizing a yoga ball or moving the recycling-bin to a goal location.

3. Representing the object

The robot's description of physical objects is a symbolic abstraction of the low level continuous experience of the robot.

3.1. Continuous system

From an experimenter's perspective, a robot and its environment can be modeled as a dynamical system:

$$\begin{aligned} x_{t+1} &= F(x_t, u_t) \\ z_t &= G(x_t) \\ u_t &= H_i(z_0, \ldots, z_t) \end{aligned} \tag{1}$$

where x_t represents the robot's state vector at time t, z_t is the raw sense vector, and u_t is the motor vector. The function F encodes state transitions while the function G encodes the observation from each state. The functions F and G represent relationships among the environment, the robot's physical state, and the information returned by its sensors, but these functions are not known to the robot itself [12].

The robot acts by selecting a control law H_i such that the dynamical system Eq. (1) moves the robot's state x closer to its goal, in the context of the current local environment. When this control law terminates, the robot selects a new control law H_j and continues onward.

The raw sensorimotor trace is a sequence of sense and motor vectors.

$$\langle z_0, u_0\rangle, \langle z_1, u_1\rangle, \ldots, \langle z_t, u_t\rangle, \ldots \tag{2}$$

3.2. Symbolic abstraction

The components of the object knowledge are represented by a tuple,

$$\mathcal{O} \equiv \langle \mathcal{T}, \mathcal{P}, \mathcal{C}, \mathcal{A}\rangle \tag{3}$$

consisting of trackers ($\mathcal{T}$), perceptual functions ($\mathcal{P}$), classes ($\mathcal{C}$), and actions ($\mathcal{A}$).

An *object*, considered as part of the agent's knowledge representation, is a hypothesized entity that accounts for a spatio-temporally coherent cluster of sensory experience. Note that the word "object", when used in this sense, does not refer to a physical thing in the external world, but to something within the agent's knowledge representation that helps it make sense of its experiences.

A *tracker* $\tau \in \mathcal{T}$ names two corresponding things:

(1) the active process that tracks a cluster of sensory experience as it evolves over time, and
(2) the symbol in the agent's knowledge representation that represents the object (i.e., the hypothesized entity that accounts for the tracked cluster).

A *perceptual function* $f \in \mathcal{P}$ is used to generate the percept $f_t(\tau)$ which represents a property of τ at time t. The percept is formed from the sensory experience by the tracker τ. Examples of simple percepts include the distance or location of a particular object at a particular time. A more complex percept is the shape of an object, which can be assembled from multiple observations over time.

For a particular perceptual function f, a *class* $\sigma_f \in \mathcal{C}$ is an implicitly defined set of percepts similar to an exemplar percept $\bar{q} = f_{t'}(\tau')$,

$$\sigma_f[\bar{q}] = \{q \mid d(q, \bar{q}) \approx 0\}, \tag{4}$$

where d is a distance function (an example is given in Eq. (13)). For example, a shape class is a set of shape percepts that are similar to a prototype shape percept. Fig. 5 shows ten shape models, which are percepts obtained from the robot's sensory experience with the ten depicted objects. These individual percepts belong to ten classes, each corresponding to percepts obtained from the same real-world object on different occasions.