Towards Robot Incremental Learning Constraints from Comparative Demonstration (Demonstration)

Rong Zhang, Shangfei Wang, Xiaoping Chen[*], Dong Yin, Shijia Chen, Min Cheng,
Yanpeng Lv, Jianmin Ji[+], Dejian Wang and Peijia Shen
University of Science and Technology of China, Hefei 230026, China
[+]The Hong Kong University of Science and Technology, Hong Kong
xpchen@ustc.edu.cn

ABSTRACT

This paper presents an attempt on incremental robot learning from demonstration. Based on previously learnt knowledge about a task in simpler situations, a robot learns to fulfill the same task properly in a more complicated situation through analyzing comparative demonstrations and extracting new knowledge, especially the constraints that the task in the new situation imposes on the robot's behaviors.

Categories and Subject Descriptors

I.2.6 [**Learning**]: Knowledge acquisition

General Terms

Experimentation

Keywords

Intelligent Robot, Learning from Demonstration

1. INTRODUCTION

In recent years, researchers have shown growing interest in Learning from Demonstration (LfD) [1], which provides a new approach to improving the abilities of robots. Most LfD methods currently concentrate on learning procedural knowledge about how to fulfill a given task. However, the same task should be fulfilled differently in different situations. A procedure for fulfilling the task in a certain situation may be improper in another one, e.g., causing harmful side-effects. For example, a robot who knows how to pick up an item in ordinary situations may not know how to avoid falling of other items in some particular situations. One solution to this problem is to decompose LfD into two parts: first learning "canonical knowledge" for ordinary (simplest, typical) situations and then learning constraints to the canonical knowledge for more and more complicated unordinary situations. Therefore, the entire learning process becomes incremental and needs less number of demonstrations.

This paper presents an effort on the approach called the Learning Constraints from Comparative Demonstration (LCfCD), in which the teacher (people) demonstrates for the task in a new unordinary situation a number of right and wrong behaviors. The robot tries to recognize the differences between the right and wrong behaviors, and extract new knowledge, especially the constraints that the task in the new situation imposes on the robot's behaviors.

2. APPROACH

LCfCD assumes that the teacher and the robot "share" a set A of primitive actions. The precondition and the effect of each primitive action $a \in A$ are known by the robot and taken to be identical for both the teacher and the robot, i.e., differences between the executions of each action by the teacher and the robot are ignored. Thus it is not required to identify the teacher's actions precisely. States of the environment are specified by a subset of P, the set of predefined predicates. For instance, in our experiment, P contains $on(X, Y)$, standing for the fact that object X is on the object Y, and $sticking\_out(D)$, for D is sticking out. $\{s_1, a_1, \ldots, s_n, a_n, s_{n+1}\}$ is called an execution sequence, where s_1 is the initial state, $a_1, \ldots, a_n$ are primitive actions, and s_{i+1} is the sequential state reached by the execution of a_i under s_i. a_n is called the end action and s_{n+1} the end state. An execution sequence and a learning label $t \in \{+, -\}$ compose a task demonstration $e = \langle h, t \rangle$, where $+$, $-$ denotes right and wrong respectively. A task demonstration labeled with $+/-$ is called a positive/negative example.

The LCfCD also assumes the robot has been equipped with a general-purpose planner and a knowledge base KB which contains previously learnt knowledge about the task, including the knowledge about the primitive actions and other background knowledge. With the planner and KB, the robot can complete the task properly in the previously known situations.

The data for LCfCD $E = \langle E^+, E^- \rangle$ is composed of a set of positive examples E^+ and a set of negative examples E^-, which are obtained by behavior identification and attitude recognition. Whence E is ready, the learning procedure of LCfCD is conducted in the following steps. (1) Difference analysis: Identify the difference set $D \subseteq P$ between the end states of execution sequences in E^+ and E^-, where D is included in every end states in E^- and none of end states

[*]Corresponding author.

Cite as: Towards Robot Incremental Learning Constraints from Comparative Demonstration (Demonstration), Zhang et al., *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 1301-1302.

in E^+. (2) Causal analysis: Extract, if any, new rules of primitive actions describing their unexpected effects observed in E. For instance, in out experiment, a new rule, R, is learnt: if there is a red can on the sticking-out end of the board and the blue can on the other end of the board is picked up, then the red can will fall. (3) Pre-condition analysis: Make out initial conditions under which D is satisfied in the end states. The result is a set of predicates I which is included in the initial state of every $e \in E^-$ and not in the initial state of any $e \in E^+$. (4) Induction: Generalize the extracted knowledge into a more general form. For instance, under some conditions, predicates such as *can*, *cup*, etc can be generalized as *small_object*, and predicates such as *red* will be ignored, meaning that color is irrelevant. After the learning phase, KB is updated with learnt rules and constraints like C: $T \wedge I \Rightarrow \mathit{not}\ D$, which states that D is prohibited if the task is T and the initial state satisfies I.

An execution sequence is extracted from a demonstration of the teacher through detection and tracking of the related objects, as described as follows. (1) Pre-processing: A median filter is used for noise reduction on the captured videos and depth information. (2) Target segmentation: to narrow the region of interest, an initial segmentation is executed by making use of the mask constructed from the depth information. Then the ultimate segmentation of target objects of concern is executed in HIS. (3) Target tracking: The directions and speed of movement are calculated from the location differences of the targets in the previous and current frames, and the most likely locations of the targets in next frame are estimated. (4) Extracting information of the states and the actions. Currently we only consider primitive actions that are easy to be distinguished. For example, pick-up and put-down can be distinguished according to the direction of movement. Meanwhile, we only consider the predefined, known objects in recognizing the environmental states. As a result, a state is extracted as a set of the predicates over these objects, where each predicate is identified by the robot's vision analyses as being true at the state.

Now the +/− label is expressed by the teacher's noding/shaking her head, respectively. To recognize them, the teacher's pupils are detected first through the following steps. (1) AdaBoost is used with Haar features to detect the teacher's face region. (2) In the same way M left-eye and N right-eye regions are detected in ($x \in [1,\ \mathit{width}/2]$, $y \in [1,\ 0.6 \times \mathit{height}]$) and ($x \in [\mathit{width}/2,\ \mathit{width}]$, $y \in [1,\ 0.6 \times \mathit{height}]$) (Fig 1). If $M = 0$ or $N = 0$, then the algorithm fails; otherwise, $M + N$ coordinates of pupils are calculated according to the proportion of eye. And there are total $M \times N$ pupil pairs. (3) Three weights are summed as the probability of each pair: $W_1 = 1 - |S_l - S_r|/(S_l + S_r)$; $W_2 = \rho < 0.8\,?-10 : \rho$, where $L = (S_l + S_r) \times 5/6$ and $\rho = 1 - |L_p - L|/(L_p + L)$; $W_3 = \theta < 0.85\,?-10 : \theta$, where $\theta = D_x/(D_x + D_y)$. The pair with the largest $W_1 + W_2 + W_3$ is selected. Here 0.8 and 0.85 are empirical values. Then the horizontal and the vertical displacement of binoculus are calculated in each two successive frames. If the horizontal is larger than the vertical, then it is shake, else it is nod. Finally, vote is used to determine the expression of the whole sequence.

3. DEMO

This demo (`http://www.wrighteagle.org/`) shows one of tests on the LCfCD approach. The robot [2] has a 6-DOF manipulator, multiple cameras and laser range finders. Also the robot has a general-purpose planner and a KB as described in last section. The robot can complete tasks of moving small objects in ordinary situations where items' falling is not considered. The purpose of the experiment is to show that: with LCfCD, the robot can learn to move objects while avoiding things falling in the designed scenario.

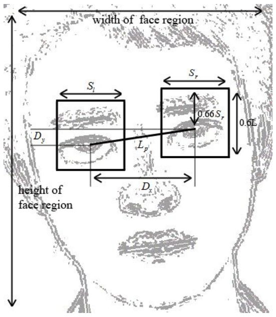

Figure 1: The proportional relation.

The teacher demonstrates a positive and a negative example of the same task, to pick up the can on the inside end of the board. In the negative example, the teacher picks up the inside can directly, causing the outside one to fall; in the positive example, the teacher puts the outside can on the table first, then to pick up the inside one. Actually, the robot can generate these two sequences with the current KB, but will always choose the wrong one because it is shorter and the current KB does not contain rules predicting the falling of items or constraints prohibiting falling of items. So this is a substantially new situation to the robot.

With LCfCD, the robot gets $D = \{can(a), red(a), on(a, b), ground(b), board(c), on(c, b)\}$, R, and C (see last section). The learned rules and constraints are generalized with background knowledge, obtaining the resulted $D' = \{on(a, b), small\_object(a), ground(b), board(c), on(c, b)\}$, as well as corresponding R' and C'; otherwise, more task demonstrations would be needed to reach the same generality.

After the learning phase, the robot is asked to pick up one of the cans. The experiment shows that, the robot can always complete the task while avoiding items falling. In addition, the robot is also asked to pick up the outside can, and she picks it up directly. This means that LCfCD does not damage the original knowledge which keeps valid.

4. ACKNOWLEDGMENTS

This work is supported by the National Hi-Tech Project of China under grant 2008AA01Z150 and the Natural Science Foundations of China under grant 60745002. We thank Fengzhen Lin, Michael Littman and Shlomo Zilberrstein for helpful discussions about this effort.

5. REFERENCES

[1] B. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009.

[2] X. Chen, J. Ji, J. Jiang, G. Jin, F. Wang, and J. Xie. Developing high-level cognitive functions for service robots. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 989–996, 2010.