# Chapter 9
# Awareness-Based Recommendation by Passively Interactive Learning:
## Toward a Probabilistic Event

**Tomohiro Yamaguchi**
*National Institute of Technology, Nara College, Japan*

**Takuma Nishimura**
*Nippon Telegraph and Telephone West Corporation, Japan*

**Shota Nagahama**
*National Institute of Technology, Nara College, Japan*

**Keiki Takadama**
*The University of Electro-Communications, Japan*

## ABSTRACT

*In artificial intelligence and robotics, one of the important issues is to design human interface. There are two issues: One is the machine-centered interaction design. Another one is the human-centered interaction design. This research aims at the latter issue. This chapter presents the interactive learning system to assist positive change in the preference of a human toward the true preference. Then evaluation of the awareness effect is discussed. The system behaves passively to reflect the human intelligence by visualizing the traces of his/her behaviors. Experimental results showed that subjects are divided into two groups, heavy users and light users, and that there are different effects between them under the same visualizing condition. They also showed that the authors' system improves the efficiency for deciding the most preferred plan for both heavy users and light users. As future research directions, a probabilistic event and its basic recommendation way are discussed.*

DOI: 10.4018/978-1-5225-5276-5.ch009

# INTRODUCTION

## Interactive Reinforcement Learning with Human

A long term goal of interactive learning system is to incorporate human to solve complex tasks. Reinforcement learning is the Standard behavior learning method for among robot, animal and human. In interactive reinforcement learning, there are two roles, a learner and a trainer. The input of a reinforcement learner as a learning goal is called a *reward*, and the output of the learner as a learning result is called a *policy*. For example, as training a dog by a human trainer, Peterson (2000, 2001) showed that clicker training is an easy way to shape new behaviors. When a dog performs a new behavior to learn, the trainer clicks the clicker as a positive reward. Pryor (2006) remarks that clicker training is a method for training an animal that uses positive reinforcement in conjunction with a clicker to mark the behavior being reinforced under behavior modification principles.

In current researches of interactive reinforcement learning, there are two approaches to support a learner by giving feedback as, whether a learning goal (reward based), or a learning result (policy based). The former approach is clicker training for the robot, in that a human trainer gives a learning goal to the robot learner. In field of robot learning, Kaplan et al. (2002) showed that interactive reinforcement learning method in that reward function denoting goal is given interactively has worked to establish the communication between a human and the pet robot AIBO. The main feature of this method is the interactive reward function setup which was fixed and build-in function in the main feature of previous reinforcement learning methods. So the user can sophisticate reinforcement learner's behavior sequences incrementally.

Ng et al. (1999) and Konidaris & Barto (2006) showed that reward shaping is the theoretical framework of such interactive reinforcement learning methods. Shaping is to accelerate the learning of complex behavior sequences. It guides learning to the main goal by adding shaping reward functions as subgoals. Previous reward shaping methods have three assumptions on reward functions as following:

- Main goal is given or known for the designer;
- Marthi (2007) remarks that subgoals are assumed as shaping rewards those are generated by potential function to the main goal;
- Ng et al. (1999) showed that shaping rewards are policy invariant, it means not affecting the optimal policy of the main goal.

However, these assumptions will not be true on interactive reinforcement learning with a non-expert end-user. Main reason is discussed by Griffith et al. (2013) that

## Related Content