Two-day seminar on

Stochastic Dynamic Programming and Temporal Difference Reinforcement Learning

Hino Campus, Tokyo Metropolitan University
（首都大学東京日野キャンパス）

October 13 and 14, 2016

10:00am–11:30am

國立台灣科技大學 助教授 水谷英二 先生 (Prof.Eiji Mizutani)

We begin with the fundamental concept of stochastic dynamic programming (DP) and Markov decision process (MDP) using Chapter 9 of a textbook below

S. Dreyfus and A. Law ``The Art and Theory of Dynamic Programming'' (Academic Press, 1977)

We then describe the basic framework of temporal difference reinforcement learning (TDRL). In particular, we consider a sequential decision-making problem in a *non-Markovian domain*, where standard DP requires a complete mathematical model; hence, a totally *model-based* procedure. By contrast, TDRL leads to a totally *model-free* procedure for seeking a best history-dependent policy.

As a particular realization, we describe actor-critic reinforcement learning with recurrent neural networks. Here, the recurrent connections (or context units) in neural networks act as an implicit form of internal state (i.e., history memory) for developing sensitivity to hidden non-Markovian dependencies, rendering the process Markovian *implicitly and automatically* in a totally model-free fashion.