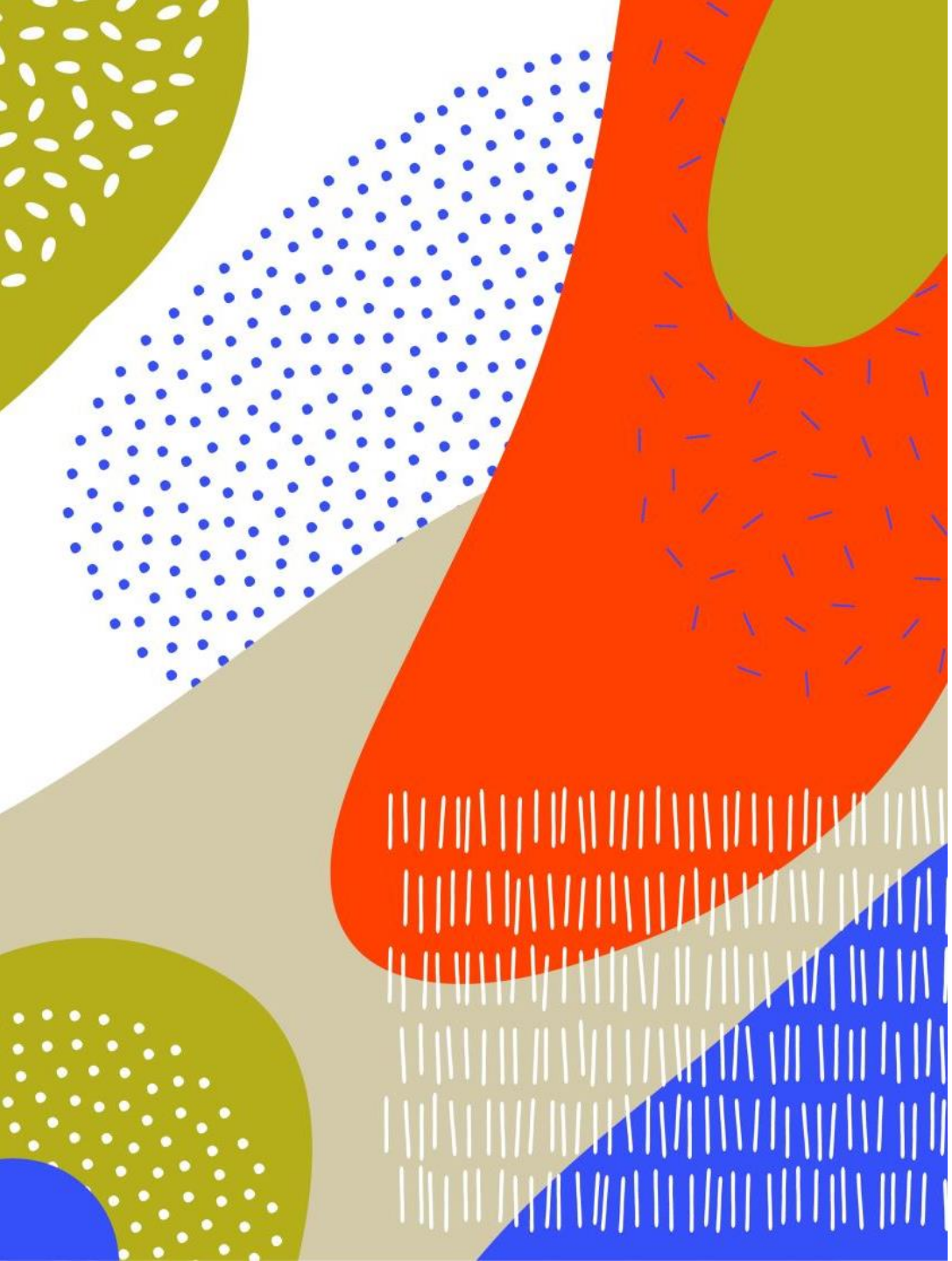

Markov Decision Process (MDP) Framework and Q-learning to solve MDPs.

by

Manoj K. Jha, Ph.D., P.E.

Professor of Practice, BITS-Pilani Hyderabad

Director of Data Science and Advanced Analytics, The Brite Group, USA

Adjunct Associate Professor of Information Technology, University of Maryland Global Campus, USA

Adjunct Professor of Civil and Environmental Engineering, George Washington University, USA

E-mail: **manoj.jha@hyderabad.bits-pilani.ac.in**

Review Material

- T2: Tom M. Mitchell: Machine Learning, The McGraw-Hill,2017
  - pp. 370-379
  - **(This coverage is better and comprehensive)**

- T1: T1: Christopher Bishop: Pattern Recognition and Machine Learning, Springer-Verlag New York Inc., 2006
- pp. 607-610

Some good references on the web

- https://builtin.com/machine-learning/markov-decision-process
- (The link above explains better and is easy to follow)

- https://neptune.ai/blog/markov-decision-process-in-reinforcement-learning

Markov Decision Process (MDP)

- Used on sequential datasets and decision making
- Used in reinforced learning
- Integer time process
- Memoryless

MDP Examples

- Real-World Applications of MDP
  - Highway Infrastructure Maintenance
  - Robot movement to perform certain action (e.g., to perform a search and rescue operation in a hostile environment; to perform a series of tasks to minimize fault and achieve perfection)

Markov Decision Process

- Jha, M.K., K. Kepaptsoglou, M. Karlaftis, and J. Abdullah (2006). A Genetic Algorithms-Based Decision Support System for Transportation Infrastructure Management in Urban Areas, in “Recent Advances in City Logistics: Proceedings of the 4th International Conference on City Logistics,” pp. 509-523, E. Taniguchi and R. Thompson (eds.), Elsevier Publishing Company, Hardbound, ISBN: 0-08-044799-6, 554 pp.

Markov decision process

The Markov Decision Process (MDP) (Figure 1) when applied to roadway maintenance allows calculation of optimal annual policies over a planning horizon (Feinberg and Schwartz 2002; Gallager 1995). In Figure 1, $\pi's$ are the condition probabilities, $X's$ are the condition states of the infrastructure elements, and $a's$ are the maintenance actions undertaken. The key features of an MDP are: (1) it is an integer time process, i.e., it assumes that maintenance actions are performed annually (integer time intervals); (2) every year a set of policy (or maintenance actions) is available of which exactly one policy is to be chosen; the cost associated with each policy is assumed to be available; (3) the transition probability leading to the condition state of the facility (i.e., the roadway segment in question) in subsequent years when a particular policy is implemented in the current year is also assumed to be exogenously available; (4) the objective is to obtain optimal annual policies over the chosen planning horizon, say T years that minimises the aggregate expected discounted maintenance cost. The mathematical theory of MDP is well described by Gallager (1995) and has been skipped here for brevity. Due to the very nature of the MDP a dynamic programming algorithm is generally applied to solve the

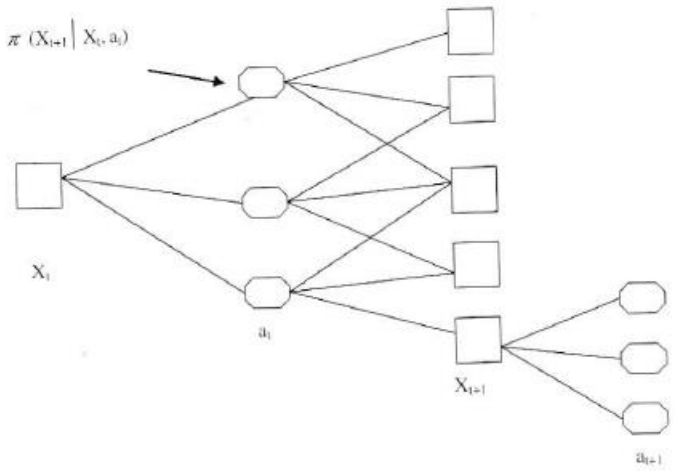

Figure 1 Decision Tree for MDP