Research Proposal

Real world is often uncertain, staring from the time taken one waits for a bus to the amount of rainfall in a given year. Planning in the face of such uncertainty is hence crucial if we are interested in designing practical systems with good performance. The framework of stochastic control serves as a useful mathematical tool to model and solve problems of planning under uncertainty. Some common instances of stochastic control problems are

- Networking: Control of Network of Queues, Traffic Control.
- Business: Inventory Control.
- Artificial Intelligence: Terrain Exploration by an autonomous Robot
- Energy: Optimal energy harvesting in Sensor Networks, Control of Smart Grids.

In general several dynamic resource management problems can be cast in the framework of stochastic control. Some of the interesting research directions related to theory and practice of stochastic control are as under.

1 Approximate Dynamic Programming

A large number of stochastic control problems occurring in practice belong to the sub-class called Markov Decision Processes (MDPs) [2]. The Markovian assumption in the system dynamics enables us to solve MDPs using the principle of Dynamic Programming (DP). Exact solution methods for MDPs are built on the idea DP are not applicable to MDPs occurring in practice which typically have large number of states.

Approximate Dynamic Programming (ADP) algorithms [6] tackle the issue of large number of states by cleverly combining approximate representations and the DP principle. While several algorithms [2] exist in literature, not all of them have guaranteed performance. An important and fundamental research question is to understand conditions under which ADP algorithms can be guaranteed to perform well in practice.

2 Reinforcement Learning

In most practical scenarios, the model information of the underlying MDP is not available. However, the system is available only in the form of a simulator or samples can be obtained via direct interaction. In such a scenario, we need to *learn* using the samples. Reinforcement Learning (RL) [9] algorithms are sample trajectory based solution methods for MDP. RL algorithms such as Q-learning [11], temporal difference learning [10] have been successful in domains such as Backgammon, elevator control etc. The aim is to develop sample efficient, stable and convergent RL algorithms for real world systems.

3 Stochastic max-Plus Systems

The class of discrete event systems where elementary components interact via 'synchronization' are called Stochastic $\max -$ Plus (SMPL) Systems [4]. Examples of such systems include railway networks, production chain, scheduling, queuing and digital systems. The inherent randomness in the different interacting components leads to the stochastic behavior. Developing approximate algorithms to derive control strategies for SMPL systems is of interest.

4 Bandit Algorithms

The classical multi-arm-bandit problem [1] is an example of trade off between exploration (trying out the arms to find the best arm) and exploitation (playing the arm currently known to be the best). Several variants of the problems include linear bandits, convex bandits, contextual bandits, bandits with complex or partial feedback have been considered in literature to model various scenarios. Exploring newer variants of the bandit problems is important.

5 Optimal Resource Allocation Problems

An interesting line of research is to empirically test and tune the performance of the state of the art ADP, RL and bandit based algorithms in various application domains. The following are the practical application domains that are of interest in the immediate future.

5.1 Pricing

Optimal pricing of resources is a common problem in various domains such as crowd-sourcing, smart grids and inventory systems. For instance, in crowd sourcing the price offered for a task affects its completion time [3]. In the case of smart grid [7], optimal pricing is key to make profits in a *Tariff* market setting. Dynamic pricing of items is important in the context of inventory management.

5.2 Sensor Network

Optimal resource allocation in wireless sensor networks has attracted attention in recent times [8, 5]. The nodes harvest energy and are self-powered. The network consists of one centralized controller that collates information from other nodes, and the center also has control over the *sleep-wake* schedules of all the other nodes. The aim here is to come up with a policy that keeps the minimum number of sensors awake per unit time while meeting the communication objective. A novel research direction would be to employ methods that exploit the topology and nature of object movement by computing the Laplacians of the associated graphs.

5.3 Expected Milestones

Quantifiable:
$2 - 3$ papers per year in top tier conferences and 1 paper per year in top tier journals.

Non- Quantifiable:

- Get involved in Research Projects relvant to IIT-Dharwad.
- Collaborate with colleagues and play vital role in the research and teaching activities at IIT-Dharwad.

References

[1] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[2] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume II. Athena Scientific, Belmont, MA, 4th edition, 2013.

[3] S. Faradani, B. Hartmann, and P. G. Ipeirotis. What's the right price? pricing tasks for finishing on time. In *Human Computation*, 2011.

[4] S. S. Farahani. Approximation methods in stochastic max-plus systems. 2012.

[5] Jason A Fuemmeler and Venugopal V Veeravalli. Energy efficient multi-object tracking in sensor networks. *Signal Processing, IEEE Transactions on*, 58(7):3742–3750, 2010.

[6] W. B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.

[7] Prashant P Reddy and Manuela M Veloso. Strategy learning for autonomous agents in smart grid markets. 2011.

[8] Vinod Sharma, Utpal Mukherji, Vinay Joseph, and Shrey Gupta. Optimal energy management policies for energy harvesting sensor nodes. *Wireless Communications, IEEE Transactions on*, 9(4):1326–1336, 2010.

[9] R. S. Sutton and A. G. Barto. *Introduction to Reinforcement Learning.* MIT Press, Cambridge, MA, USA, 1st edition, 1998.

[10] John N. Tsitsiklis and Benjamin Van Roy. An analysis of temporal-difference learning with function approximation. 1997.

[11] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.