Statement of Purpose

Robotics Institute — Ph.D. Applicant
Carnegie Mellon University — Subhojyoti Mukherjee [CV]

I want to pursue Ph.D. in Computer Science and I aspire to become a professor in this field. My research interests span the areas of *Machine learning, Reinforcement learning, Online Learning, Multi-armed bandits, Applied Probability, Optimization*. I completed M.S (Research) in Computer Science on July 2018 from **Indian Institute of Technology Madras** under the supervision of Dr. Balaraman Ravindran and Dr. Nandan Sudarsanam. I am currently a first semester Ph.D. candidate at **University of Massachusetts Amherst** from Fall 2018 but I am looking for Ph.D opportunities in Carnegie Mellon University as my recruiting faculty Dr. Akshay Krishnamurthy has left for Microsoft Research and there are no other faculty at the current institute whose research interest matches with mine.

1 Completed/Ongoing Research

I have mainly focused on theoretical research in Reinforcement learning (RL). To briefly describe, an RL agent employs a sequentially adaptive strategy to select actions based on some environmental feedback that maximizes some long-term performance metric. Within the RL framework, Multi-armed Bandits address one of the fundamental challenges of learning in general called the exploration-exploitation dilemma which tries to answer the following question: "When should a learning algorithm stop exploring alternative actions and focus on the action that has yielded the maximum payoff till now?" This trade-off is a fundamental challenge in many practical and industrial level problems such as item recommendation to new users for which no prior information exist; administration of medical treatments to suffering patients; design of a speech and dialogue system in Natural Language Processing, intelligent tutoring system, etc. Some of my research works which got published in premiere conferences (or is ongoing) which addresses these issues are mentioned below.

1. Variance aware Bandits: An important problem which arises in mobile channel recommendation and online item recommendation where there is a separation between testing and deployment phases is how to employ these bandits in a pure-exploration setting. In the pure-exploration Thresholding Bandits (**Mukherjee, Subhojyoti** et al., 2017) we proposed an algorithm which finds the best set of actions with qualities above a particular threshold. The challenge lies in the observation that this best set can be very large and moreover that the candidate set of actions can be exponentially large. Our novel algorithm solves this problem by estimating the mean and the *variance* (hence being variance-aware) of the actions to determine their qualities, thereby significantly reducing the exploration. We also provide theoretical bounds for the maximum possible loss suffered by our method. This work was published in **International Joint Conference on Artificial Intelligence 2017**.

My next work focused on a more fundamental problem regarding exploration-exploitation dilemma in which there has been a theoretical gap in the lower and upper bound on the loss of a class of strategies called *variance aware elimination* algorithms. These algorithms eliminate actions which are deemed to be sub-par depending on their qualities based on estimated variances. Our proposed algorithm called Efficient-UCB-Variance (**Mukherjee, Subhojyoti** et al., 2018) falls under this class, and we established that the maximum possible loss (regret upper bound) of this strategy matches the lower bound for all classes of strategies in all the possible environments satisfying certain properties. This works have been accepted in **Association for Advancement of Artificial Intelligence 2018**.

2. Changepoint detection and piecewise i.i.d Bandits: Applying Bandit strategies in environments where the quality of all the actions change abruptly and simultaneously has proved to be quite challenging. This is quite common in *medical domain* where the behavior of drug-resistant bacteria may change abruptly prompting a different response to all treatments (actions). We proposed two novel algorithms called UCB-CPD and ImpCPD for this setting with interesting theoretical guarantees and closed some gaps in the existing bounds. This work was jointly done with Dr. Odalric Maillard while I was an intern at INRIA, SequeL Lab, Lille and is currently under review in **International Conference on Artificial Intelligence and Statistics 2019**.

3. Ranking and Latent Bandits: Online ranking of items personalized for each user based on their historical preference is an active area of research. By leveraging its low rank structure previous works have mainly focused on its

approximate reconstruction using stochastic observations that require many assumptions and costly matrix operations. We mainly focused on intelligently exploring a partially observed user-item matrix without reconstructing it by leveraging the prior knowledge of the rank of the matrix. This is an ongoing joint work with Dr. Branislav Kveton, and Dr.Anup Rao while I was an intern at Adobe Research, San Jose which we are planning to publish soon.

These works and more fundamental insights into several aspects of Bandits, RL and Learning theory have been captured in my MS thesis which was accepted by my technical committee at IIT Madras without any major revision.

2 Future Directions

Working on theoretical research in RL setting now allows me to expand my research frontier in a variety of new but related research areas including but not limited to statistical learning theory and optimization and to explore problems that bridge the gap between learning theory and empiricism. A few of the interesting research problems that I intend to work are,

1. Off-policy Reinforcement Learning: Often the complete knowledge of the world is not available to the learner forcing it to adopt different strategies. Examples include maze navigation by a robot maze using faulty sensors, medical treatments with only a partial knowledge about the true state of the patient (as it is requires information to an atomic level), and user preference recommender systems without cumbersome exploration strategies of complete user logs. Partial observability is a broad area of research in RL, and providing theoretical guarantees is hard due to the high variance in the performance of various strategies given only a limited state-space exposure. I can draw from my previous experience as this problem boils down to intelligent exploration for the reduction of performance variance and to provide theoretical guarantees

2. Safe Reinforcement Learning:I am interested in safe RL which also poses some of the same challenges as above. For example, in medical domain, not only do we have to work in a partial observable environment but we should also give confidence guarantees for the proposed action (medical treatment) since this gives confidence to the medical practitioner in a high-risk environment where faulty actions might cause death. This is again directly related to my previous experience as this requires a sound knowledge of concentration inequalities to give the confidence on an action taken. Safe RL is also of extreme importance in areas of robotics as well, for example in autonomous driving, and motion planning and control.

3. High Dimensional Reinforcement Learning: There has been several recent successes of RL in large state spaces such as the game Go and autonomous driving, wherein the success has come by employing powerful nonlinear function approximation techniques like Neural Networks. But sound theoretical guarantees for RL only exist for small state spaces using linear function approximation techniques. I am interested in bridging this gap between the theoretical and empirical performances of RL as this will directly benefit some of the challenges mentioned before, and I can draw my knowledge from concentration measure theory and RL to contribute to this problem.

4. Optimal Design and Active Learning: Many supervised learning tasks operate only on limited labeled data, and hence, the learning models result in high uncertainty. Moreover in certain tasks, it is costly to run experiments, and hence, the learner needs to carefully pick subsets of data to minimize model uncertainty. This can be handled in several ways including active learning, online learning, and bandit feedback. This is a combinatorial optimization problem having many open-ended questions and requiring strong theoretical guarantees, to which I can contribute from my past experience.

3 Program and Faculty of Interest

I believe that a doctoral degree from Carnegie Mellon University will propel me a long way in my goal to become a professor in Computer Science. Its broad reach and in-depth courses on a variety of subjects ranging from practical to theoretical applications in Computer Science will help me in my endeavor. There are several professors at Carnegie Mellon University such as *Dr. Pradeep Ravikumar* and *Dr. Aaditya Ramdas* whose works/projects are especially appealing to me. Dr. Pradeep Ravikumar works in theoretical machine learning and its applications to high dimensional real-life data-sets which is an area I am interested to work in. Specifically, I want to work with him in the theoret-

ical areas of learning theory and find common solutions where my expertise can be applied. Also, I am interested to work with Dr. Aaditya Ramdas in the area of online learning and optimization. I want to collaborate with him in several of his areas of interest such as automated discovery and efficient data processing in applied sciences including health-sciences and robotics, where there are several bandit variants which can be applied.

References

Mukherjee, Subhojyoti; Kolar Purushothama, N.; Sudarsanam, N.; and Ravindran, B. 2017. Thresholding bandits with augmented ucb. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2515–2521.

Mukherjee, Subhojyoti; Kolar Purushothama, N.; Sudarsanam, N.; and Ravindran, B. 2018. Efficient-UCBV: An almost optimal algorithm using variance estimates. In *Proceedings of the 32nd Association for the Advancement of Artificial Intelligence*.