

Statement of Purpose

Language Technologies Institute
Carnegie Mellon University

Ph.D. Applicant
Subhojyoti Mukherjee [[CV](#)]

I want to pursue Ph.D. in Computer Science from the Language and Technologies Institute 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*, and *Online 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 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, and intelligent tutoring system. Some of my research works which got published in premier conferences (or are ongoing) which address 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 achieved 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 bounds 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 work has 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, France, 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 pref-

erence 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](#) which was started while I was an intern at Adobe Research, San Jose, and we are planning to publish it 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 (LTI)

Working on theoretical research in RL setting allows me to expand my research frontier in a variety of new but related research areas including but not limited to statistical learning theory, online optimization, and online personalized recommendation and ranking of items. Furthermore, there are areas in Natural Language Processing which will benefit from my knowledge of Reinforcement Learning. A few of the interesting research problems that I intend to work are,

1. Intelligent Tutoring Systems: The task of designing an intelligent question-answer based tutoring system that adapts to the changing nature of the student/viewer is a challenging task. One of the main challenges is that if the system continuously provides either a very trivial task or a very challenging task (questions), then the student might lose interest and leave the task. So, the system must give a sequence of tasks that closely adapts to the learning curve of the student, varying the difficulty level slowly, where the feedback is also in form of text from the student. This property can be captured by, for instance, a recurrent neural network which learns its features based on sequential data, which in this case is represented by the series of tasks given to the student in words. The adaptive learning behavior of this problem falls under the exploration-exploitation dilemma of RL to which I can contribute.

2. Prediction from Patient EHR data: This is an area that is not only challenging but directly sits across two of my most passionate areas of research, Safe Reinforcement Learning and learning text representations. Note that a typical EHR patient's health record consists of a series of patient-physician interactions resulting in multiple electronic prescriptions (text) for the patient and their feedback. The first challenge which is slightly easier to solve is to find the correlation of diseases and administered/suggested drugs in one such prescription. A more challenging task is to connect many such prescriptions generated sequentially over time and provide a long term prediction with high guarantee. Again, I can draw from my experience in RL and learning theory to contribute to this problem. Currently, one of my research projects focuses on predictions from the MIMIC 3 dataset containing such EHR of patients.

3. Text-based Game playing: In text-based games, the main challenge lies in the ability to learn optimal policies in systems where the action space is defined by sentences in natural language, continuous state space and games with multiple endings and rewards. These feedbacks allow us to employ reinforcement learning techniques and jointly learn text representations and control policies. Again, the main challenge lies in constructing text representations that captures the long term correlations of actions and employing RL to modify the learned policy.

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 theoretical 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*.