RESEARCH STATEMENT

YASH SATSANGI

Building autonomous agents that are capable of intelligent perception and exhibit rational behavior is one of the fundamental challenges in the field of Artificial Intelligence. Akin to this objective, my research goal is to design and analyse novel algorithms that can help an agent learn from its past experience and behave rationally in complex environments. To develop such an 'intelligent' agent my research combines methods from diverse disciplines, namely, reinforcement learning, submodular optimization, probably approximately correct (PAC) learning and decision-theoretic planning.

My research methodology aims to identify key challenges of a learning or planning problem, formulate it soundly with help of the existing literature, develop efficient algorithms with firm theoretical guarantees and analyze the nature and performance of these algorithms in practical settings. This is reflected in my current and prior research contributions which present a well-balanced blend of new algorithms, theoretical results and empirical analysis. During the last three years, my focus was on active perception for tracking people in multi-sensor systems. This research led to multiple important theoretical and empirical results for long term planning in partially observable Markov decision processes (POMDPs) [7], submodular optimization [8] and real-time tracking systems [10]. Application of these methods has made it possible to scale tracking algorithms to large multi-camera networks that are typically employed in shopping malls and airports [9, 10]. This document further details my past research experience, research interests and future ambitions.

1. Active Perception and Sensor Selection

One of the key challenges in the design of multi-sensor surveillance systems is the efficient allocation of scarce resources such as manpower, bandwidth, computational power or energy [13]. Surveillance is an example of an active perception [1] problem where an agent must select k out of the n available cameras to reduce the uncertainty over the state of the world. Typically, reducing uncertainty is only a means to an end, however, in an active perception task reducing uncertainty is an aim in itself.

In my dissertation I propose multiple methods that tackle the challenges of active perception problems and in turn enable an agent to take actions to reduce its uncertainty. The active perception problem is modeled as a POMDP [11, 3], a natural decision-theoretic model for such problems, that allows an agent to compute long-term strategies (or policies) that maximize a reward, which encodes the objective of the agent.

Rewarding Certainty in POMDP. My prior work [9] addresses the challenge of rewarding a POMDP agent for reducing uncertainty by showing the equivalence of two variations on the existing POMDP framework, namely, ρPOMDP and POMDP with Information Rewards (POMDP-IR). This equivalence leads to a general insight that if the action space of a POMDP agent is augmented with prediction actions that give the agent the choice to predict the true state of the world with certain confidence and if the agent is rewarded for correctly predicting the state then this reward function is approximately the same as rewarding the agent for reducing entropy over the hidden state. This insight lets us express the entropy-based reward of an active perception POMDP agent in terms of the state-action pairs, which is imperative for employing traditional POMDP solvers. Our experiments show that for tracking people in a hallway entropy-based rewards have a 14% advantage over otherwise popular state-based rewards. For tracking people in complex environments such as shopping mall entropy-based reward outperform state-based reward functions by 35%.

Deep Anticipatory Networks. Based on the insight that the entropy based reward can be expressed in terms of the state and the prediction actions, I am currently working on deep anticipatory network (DAN), a deep neural network architecture that enables an agent to take actions to reduce its uncertainty without having to maintain an explicit representation or model of the world. A DAN agent simultaneously learns deep representations of its belief and a policy that helps the agent to make correct predictions about the state of the world. Our experiments show that DAN is able to track people in a simulated setting while observing only one-eighth of the original scene even in the presence of sensor noise.

2. Submodularity and Information gathering

One of the key challenges of active perception problems is the combinatorial action space. Traditionally, scaling in the state space of a POMDP is possible, however, combinatorial action spaces still remains a challenge for decision making. My previous research shows how submodularity, a property of set functions that formalizes the notion of diminishing return, can be exploited for scaling in the action space of the active perception POMDP.

Scaling in Action Space. Our paper on *submodular value functions for POMDPs* [7] proposes a new planning method, greedy point-based value iteration (PBVI), which addresses the challenge of scalability in the action space of a POMDP agent. Greedy PBVI scales better in the action space of a POMDP agent because it substitutes full maximization with greedy maximization in the Bellman optimality equation. As a justification of this approximation, we extend the result of Nemhauser et al. [6] on greedy maximization for submodular functions to long-term planning which involves multiple applications of greedy maximization. Moreover, we establish the sufficient conditions for submodularity of the value function of a POMDP. Our empirical analysis shows that greedy PBVI scales to large sensor selection problems and results in same tracking performance as state-of-the-art POMDP planners but at a fraction of computational cost (faster at least by a factor of 8!). Effectively, this leads to better resource allocation in large camera networks with resource constraints.

PAC Greedy Maximization. To scale greedy maximization of submodular functions to settings where it is not possible to exactly evaluate the submodular functions we propose a probably approximately correct (PAC) [12] version of greedy maximization. PAC greedy maximization uses upper and lower confidence bounds to maximize the submodular function. Moreover, we propose new computationally cheap upper and lower confidence bounds on information gain which can be combined with PAC greedy maximization to perform greedy maximization of information gain in large real-life settings. Effectively this leads to a fast and accurate submodular maximization algorithm that thrives on the integration of the optimization and evaluation of the submodular function that is to be maximized. Experiments on a real-life dataset collected from a shopping mall show that PAC greedy maximization scales to extremely large problems (problem of size 22500 $\times$ 6840 $\times$ 22500) without a big compromise on the performance.

Sampling in Action Space. To scale tracking systems to ultra high resolution (bird-eye's view) images taken by resource constrained systems, example, a drone, we present PartiMax, a tracking system that can intelligently allocate its own resources. PartiMax works on the same principle as particle filters and exploits the information in the particle beliefs maintained by the particle filter to sample locations in an image with highest probability of detecting a person. Building upon this sampling scheme PartiMax performs stochastic greedy maximization of a submodular utility function to select k out of n locations in real-time to apply a trained person detector on to track people. Furthermore, PartiMax performs this maximization with an error bound that is independent of the size of the problem. Experiments on a real-world dataset show that PartiMax can select 40 locations out of 7200 (thus processing less than 5% of original image) in less that 0.09 seconds while retaining 80% of tracking performance for multi-person tracking. In essence, PartiMax enables tracking people in large images (irrespective of their size) without any need of big computational/energy resources.

3. Future Ambitions

As illustrated by my prior work, I am interested in designing scalable real-time systems that are not just accurate but also resource efficient and that come with meaningful theoretical guarantees. A key question that I am interested in is 'how to make an agent reason about its own uncertainty and how to teach an agent to take actions to reduce it?' I wish to continue pursuing the complete answer to this question. I wish to divide this investigatory process in the following three divisions:

3.1. **Human Computer Collaboration.** We formulate active perception as a task where the end goal of an agent is to reduce uncertainty in its belief. While perception is arguably always made to aid decision-making, in an active perception problem that decision is made by another agent such as a human, that is not modelled as part of the active perception agent.

This modularity in perception and decision-making is the key to enhance the intelligence of a human decision maker for sensitive and complex problems, example, surveillance, detecting cancer cells in images, blocking extremist/sexual content on public websites, etc. For example, in the surveillance task the agent might display a suspicious activity, but only the human users of the system may decide how to react to it. The research question I am interested to address here is 'Can an agent learn to gather relevant information to aid a certain decision making process directly from its experience with the decision maker?' Based on deep Q networks [5] we propose DANs that provide one such framework where an active perception agent can be rewarded for collecting information to aid a human user in decision making directly from the behaviour of the human. I wish to extend this potential use of DAN, combined with other ideas from my research to make it a practical solution for sensitive problems such as surveillance, healthcare, finance, etc.

On the same note but in a different direction the human and computer relationship can also be that of a teacher and a student. Questions like 'when to ask for the input from a human to help an agent learn an optimal behaviour' already are being investigated with the help of reinforcement learning methods. Submodularity already plays a key role in answering such questions and thus, naturally, I am excited to pursue research to answer such questions.

3.2. **Reinforcement Learning for Active Perception.** Taking actions to reduce uncertainty comes naturally to humans, example, questions like 'where to focus', 'what to remember', 'how to navigate' etc are naturally easy for humans to answer. Autonomous agents, however, still struggle at these tasks. Thus, a straightforward direction of research is general active perceptions tasks, such as motion planning for information gathering, attention mechanism, control of memory, viewpoint selection etc. (Deep) reinforcement learning methods exactly tackle these question by enabling an agent to learn to take actions that maximize the expected cumulative reward. Combining this with my past research I intend to tackle the challenges of uncertainty representation and approximation, combinatorial action space, long-term reasoning, etc to enable an agent to learn to actively perceive.

3.2.1. *Entropy Estimation.* Entropy estimation is one of the central topics in many fields including statistics and machine learning. The existing approaches address the question of accurate entropy estimation. Estimating entropy for decision making is a fundamentally different question that involves estimating the difference in entropies of two probability distribution in real-time. My research takes the first step towards addressing this challenge by integrating the optimization algorithm with the entropy estimation algorithm to fasten the decision making. Decision making for information gain is an exciting line of research because of the numerous theoretical and practical results that are possible here. Key contributions in this direction can be identifying efficient representations for uncertainty, efficient and accurate surrogate functions (for example weighted sum of coverage function combined with a notion of diversity is observed to substitute quite well), and customized or black-box (Bayesian) optimization algorithms for information gain.

3.2.2. *Learning Active Perception.* DAN provides a deep reinforcement learning framework for an agent to learn deep representations and optimal behaviour for active perception. DAN can learn the optimal behaviour directly from a dataset and without complex modelling assumptions. Enabling DANs to be truly end-to-end is one of the most straightforward and natural direction of future work for my research.

This research ties with the recent work on learning active learning strategies. Active learning enables an agent to observe the most informative data points in a supervised learning setting. Reinforcement learning can help an agent learn an active learning strategy from its experience. However, an interesting question to answer here is that 'Under what conditions can an agent transfer an active learning strategy that the agent learned from its experience with one domain/dataset to another?'. A natural application of such a methodology is data efficient machine learning.

3.3. **Submodularity.** Submodular function optimization is an exciting research field with deep theoretical repercussions and practical applications. A common theme of my research is the integration of the optimization and evaluation of submodular function. For task that involve real time combinatorial/submodular optimization such integrations are of critical importance to balance the accuracy vs cost trade-off. I hope to keep this a frequently occurring theme of my research.

Another interesting setting for submodular optimization that is common in practice is the sequential maximization of submodular functions or 'optimization of submodular functions that are sampled from a unknown probability distribution'. Sensor selection or active perception tasks are practical applications of sequential submodular optimization. In future I plan to explore how the sequential nature of these tasks can be exploited further for accurate and faster approximate submodular optimization.

Finally, learning submodular functions is known to be a tough problem, however, it can be relaxed by either making domain specific assumption or by learning only the relevant properties of it, for example, only its maximum or minimum. Tied to this problem, I am interested in developing sampling methods that can sample from a submodular point processes [2]. I hope to borrow from sampling methods for other point processes such as determinantal [4] and hawkes process to learn more about this problem. While my experience with this area of research is limited, I am definitely looking forward to it.

References

[1] Ruzena Bajcsy, Yiannis Aloimonos, and John K Tsotsos. Revisiting active perception. *arXiv preprint arXiv:1603.02729*, 2016.

[2] Rishabh K Iyer and Jeff A Bilmes. Submodular point processes with applications to machine learning. In *AISTATS*, 2015.

[3] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, May 1998.

[4] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *arXiv preprint arXiv:1207.6083*, 2012.

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[6] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functionsi. *Mathematical Programming*, 14(1):265–294, 1978.

[7] Yash Satsangi, Shimon Whiteson, and Frans Oliehoek. Exploiting submodular value functions for faster dynamic sensor selection. In *AAAI*, pages 3356–3363, January 2015.

[8] Yash Satsangi, Shimon Whiteson, and Frans Oliehoek. PAC greedy maximization with efficient bounds on information gain for sensor selection. In *IJCAI*, pages 3220–3227, July 2016.

[9] Yash Satsangi, Shimon Whiteson, Frans Oliehoek, and Matthijs Spaan. Exploiting submodularity for scaling up active perception. *Autonomous Robots*, 2017.

[10] Yash Satsangi, Shimon Whiteson, Frans A. Oliehoek, and Henri Bouma. Real time resource allocation for tracking systems. In *UAI*, 2017.

[11] Edward J. Sondik. *The Optimal Control of Partially Observable Markov Processes*. PhD thesis, Stanford University, United States – California, 1971.

[12] Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[13] Jason L Williams, John W Fisher III, and Alan S Willsky. Sensor management for multiple target tracking with heterogeneous sensor models. In *Proc. SPIE*, volume 6235, 2006.