Sample article to present `elsarticle` class[☆]

Author One[a,b,1,*], Author Two[c], Author Three[a,c]

[a] *Address One*
[b] *Address Two*
[c] *Some University*

Abstract

Text of abstract. Text of abstract. Text of abstract. Text of abstract. Text of abstract.

Keywords: foraging, active learning, template or
2008 MSC: code, code (2000 is the default)

1. Introduction

1. Why did I do the work? Robots exploring the world right now either depend highly on their controllers to give them objectives or they are planning with some global knowledge. Operating in these kinds of conditions puts constraints on robot exploration opterations by relying on either humans to make decisions or a significant amount of scouting.
   Relying on humans to make decisions means that remote operators require considerable bandwidth to acquire sufficient situational awareness. Conducting sufficient reconnaissance to make good decisions often obviates the need to send a robotic agent.
   What is lacking in the literature are robots that make decisions about what to investigate *in situ* without reliance on humans and without necessarily having global knowledge.
2. What were the central motivations and hypotheses? Animals, e.g. human geologists, make decisions about investigating phenomena in the world without necessarily having access to high resolution satellite imagery. Despite this lack they are able to chose between sampling from materials in front of them and moving on to determine more profitable sampling locations.

[☆] This is only an example

[*] I am corresponding author
Email addresses: `author.one@mail.com` (Author One), `author.two@mail.com` (Author Two), `author.three@mail.com` (Author Three)
URL: `author-one-homepage.com` (Author One)

[1] I also want to inform about...

[2] Small city

While these decisions may not be globally optimal they do demonstrate an ability that is lacking from exploration robots: to make decisions to stop and engage with the environment or to continue travelling in the hopes of finding more informative sampling locations.

2. Background

Who else has done what?

- Work in the 1970's about foraging. About making value judgements. - Evidence that humans make (approximately) rational decisions (we over and under estimate low and high probabilities)

- Design of experiments has led to (amongst other things) multi-armed bandit models of sequential experiment design. - See also maximum entropy sampling - See also mutual information sampling. - Also consider active learning solutions (they all end up being the same anyway)

- Robotics research has made robots that conduct exploration, but the only ones that make decisions about whether to investigate something or not do one of three things: 1. match templates. 2. seek improbable things. 3. Engage in opportunistic science - they do something if they have the time. They don't override human mission objectives.

1 and 2 say nothing about the information content of the material under investigation. 3 does not have the level of autonomy that we need for truly long-term or remote operations.

How?

What have we previously done: - D.R. Thompson's work - Only looking at satellite imagery. Good but not sufficient. - Trey's work - Using POMDPs not scalable to a planet. - Mine. Where does my previous work fall short? - Not bayesian (not a big deal?) - Still has the problem on knowing the number of sampling opportunities remaining.

3. BACKGROUND

Previous approaches to planetary scale science autonomy fall down in two respects. Firstly, these approaches model scientific exploration as a standard exploration/exploitation problem. A model that does not necessarily hold for planetary exploration. Secondly, they do not use the output of the scientific measurements to improve how the robots select between sampling actions. For stationary processes experiment design dictates that the optimal set of experiments can be determined without ever knowing the results of those experiments [**?**].

3.1. Sequential Action Selection

Sequential experiment selection, a type of active learning, is addressed in the multi-armed bandit literature. The multi-armed bandit was introduced in [**?**] as a means of sequentially selecting which experiments to conduct with a limited budget. In Robbins' work [**?**] selecting experiments is modelled on determining the payouts of one-armed bandit machines – each machine represents a different experiment. The player has a fixed

sampling budget and has to sequentially choose which machine to play, trading off exploiting the expected rewards for the different arms and exploring the different arms learning more accurately the payouts of those arms.

Lai *et al.* [**?**] introduced the Upper Confidence Bound (UCB) rule which values sampling opportunities with the sum of the expected reward for a sampling opportunity and a term that tries to balance the samples amongst all types of sampling opportunities.

$$Value = \mathbb{E}\left[R_i\right] + \sqrt{\frac{2 \ln t_i}{T}}$$

Where R_i is the reward for sampling opportunity i, t_i is the number of times i has been sampled, and T is the total number of samples distributed. Work on proving the bounds of this algorithm has been continued by Agarawal [**?**] and Auer and Ortner[**?**].

Other approaches to the bandit problem use reward plus the uncertainty of that reward to indicate value. We see this in the work of Burnetas and Katehakis [**?**] and Auer [**?**]. This is a sentiment seen in other work, like the optimistic planners of Jurgen Schmidhuber's group [**?** **?** **?** **?**]. They choose actions that maximize the expected information gain with respect to some model they are learning. The most valuable actions are the ones that result in the greatest shift in the distribution the learner is building.

Balcan [**?**] presents a method for learning classifiers by requesting samples from the input space with the greatest classification error. Classification error and uncertainty in function value are fungible quantities in this case. An analogy can be drawn between the classifiers used in [**?**] and the bandit arms used by Auer and Ortner[**?**].

Thompson and Wettergreen [**?**] maximize diversity of collected samples by using mutual information sampling. This approach ensures diversity in the collected sample set, an act that reduces uncertainty in the input space of a function. Neither mutual information nor maximum entropy sampling methods, when used with stationary Gaussian processes, take into account the dependent variable when selecting samples.

Sequential experiment selection values actions by a combination of reward and uncertainty in that reward. Since the mission of exploration is learning the reward is the reduction in uncertainty by taking actions. Seeking uncertainty is a useful way to value options presented to a learning agent but it does not address the explorer's problem of either giving up on a sampling opportunity or searching for better opportunities. Further it is not guaranteed that sampling opportunities can be accessed at no cost, an assumption commonly made when querying an oracle.

3.2. Exploration as Foraging

Active learning assumes an oracle and as such does not map well to exploration in unknown environments. In approaches like those of Robbins [**?**] or Balcan [**?**] the agent conducting experiments has at any time the opportunity to sample random variable they are characterizing. This is not the case in planetary exploration, we can only sample from those random variables that are present as robots follow their trajectories. The inaccuracy of the oracle model has been previously identified by Donmez and Carbonell [**?**].

Foraging theory provides a way to make the decision to stay or to go withouth knowledge of future opportunities. This stands in contrast to the standard exploration/exploitation problem choosing from known sampling opportunities.

Optimal foraging strategies devised by Charnov [**?**] describe how predators hunt in different geographic regions with different levels of resources. Animals make the decision to forage by comparing the value of the options it has in front of it to the expected value of what it may obtain by searching for better options [**?**], less the cost of conducting a search.

Kolling *et al* [**?**] found that humans make foraging decisions based on the arithmetic mean of the estimated values of the options they are presented with and the options that remain in the surrounding environment. From foraging literature we learn to compute the value of searching in an environment by taking the arithmetic mean of what is thought to be in that environment. The decision rule to stay or leave is a comparison between the value of the current opportunity and the expected value of the environment.

Optimal foragers considering three things when choosing to leave a resource: Expected value of the current opportunity, the expected value of the rest of the environment, and the cost of searching for new opportunities [**?**],[**?**]. To adapt foraging to exploration we need to answer the question: What is the value of an option presented to the explorer? To answer that question we look to active learning.

Previous work by the authors [**?**] addressed the problem of exploring along a transect by employing techniques from Foraging Theory. However that work did not address the fact that the sampler had a limited budget. Agents in that work did not expend all of their samples for large budgets, a problem this research addresses.

The strategy presented by Ferri *et al.* made a comparison between the perceived value of the available sampling opportunity and an arbitrary function of the remaining number of sampling opportunities [**?**]. The value of a sampling opportunity was determined by a fixed threshold and the proclivity for spending samples was likewise determined by an arbitrary constant. In contrast this work employs an information theoretic measure of opportunity value and a principled measure to determine when to expend a sample.

The prior work yields two observations. Firstly, foraging, a better model for planetary exploration, requires a measure of value of the sampling opportunities available to the exploring agent. Secondly, active learning uses uncertainty – in both input and output space of a function – to value potential exploration opportunities. What follows next is a method for exploring that reflects the limitations of a planetary setting and incorporates the result of sampling operations into decision making processes.

4. BACKGROUND

Previous approaches to planetary scale science autonomy fall down in two respects. Firstly, these approaches model scientific exploration as a standard exploration/exploitation problem. A model that does not necessarily hold for planetary exploration. Secondly, they do not use the output of the scientific measurements to improve how the robots select between sampling actions. For stationary processes Bayesian experiment design dictates that the optimal set of experiments can be determined without ever knowing the results of

those experiments [**?**]. However real world quantities are not necessarily stationarity and they may not even obey a function.

4.1. Foraging as Exploration

The exploration/exploitation problem asks the question: Is an agent rewarded better by exploiting already acquired knowledge or by exploring different options and improving that knowledge? The multi-armed bandit [**?**] was introduced to address the exploration/exploitation trade-off with a limited sampling budget. Multi-armed bandits model a fixed list of experiments as different slot machines each with their own random payout. An arm of a bandit is a metaphor for a random variable and the reward for playing that arm reveals information about that random variable. A shortcoming of the multi-armed bandit approach is that it assumes that at any given time all random variables are known and are available to conduct.

Active learning assumes an oracle and as such does not map well to exploration in unknown environments. In approaches like those of Robbins [**?**] or Balcan [**?**] the agent conducting experiments has at any time the opportunity to sample random variable they are characterizing. This is not the case in planetary exploration, we can only sample from those random variables that are present as robots follow their trajectories. The inaccuracy of the oracle model has been previously identified by Donmez and Carbonell [**?**].

Foraging theory provides an answer to the question of whether to stay or to go in the face of unknown future opportunities. This stands in contrast to the standard exploration/exploitation problem choosing from known sampling opportunities.

Optimal foraging strategies devised by Charnov [**?**] describe how predators hunt in different geographic regions with different levels of resources. Animals make the decision to forage by comparing the value of the options it has in front of it to the expected value of what it may obtain by searching for better options [**?**], less the cost of conducting a search. The distinction between exploration/exploitation and forage/engage is determined by two things: the recognition that there is not always a choice of what to explore and the realisation that the choice is between what is available and what may yet be encountered.

Kolling *et al* [**?**] found that humans make foraging decisions based on the arithmetic mean of the estimated values of the options they are presented with and the options that remain in the surrounding environment. From foraging literature we learn to compute the value of searching in an environment by taking the arithmetic mean of what is thought to be in that environment. The decision rule to stay or leave is a very simple comparison between the value of the current opportunity and the value of the environment.

Optimal foragers considering three things when choosing to leave a resource: Expected value of the current opportunity, the expected value of the rest of the environment, and the cost of searching for new opportunities [**?**],[**?**]. To adapt foraging to exploration we need to answer the question: What is the value of an option presented to the explorer? To answer that question we look to active learning.

4.2. Active learning

In active learning agents get to choose examples in order to resolve uncertainty or inaccuracy in models they are learning. An early version of active learning is the multi-armed bandit problem. The k-armed bandit was introduced in [**?**] as a means of sequentially selecting which experiments to conduct. In Robbins' work [**?**] selecting which experiment to conduct next is modelled on determining the payouts of one-armed bandit machines, where each machine represents a different experiment. The player has a fixed sampling budget and has to sequentially choose which machine to play, trading off exploiting the expected rewards for the different arms and exploring the different arms learning more accurately the payouts of those arms.

Lai *et al.* [**?**] introduced the Upper Confidence Bound (UCB) rule which values sampling opportunities with the sum of the expected reward for a sampling opportunity and a term that tries to balance the samples amongst all types of sampling opportunities.

$$Value = \mathbb{E}\left[R_i\right] + \sqrt{\frac{2 \ln t_i}{T}}$$

Where R_i is the reward for sampling opportunity i, t_i is the number of times i has been sampled, and T is the total number of samples distributed. Work on proving the bounds of this algorithm has been continued by Agarawal [**?**] and Auer and Ortner[**?**].

Other approaches to the bandit problem use reward plus the uncertainty of that reward to indicate value. We see this in the work of Burnetas and Katehakis [**?**], and Auer [**?**]. This is a sentiment seen in other work, like the optimistic planners of Jurgen Schmidhuber's group [**?** **?** **?** **?**]. They choose actions that maximize the expected information gain with respect to some model they are learning. The most valuable actions are the ones that result in the greatest shift in the distribution the learner is building.

Balcan [**?**] presents a method for learning classifiers by requesting samples in the input space of the function where the classification error is the greatest. Classification error and uncertainty in function value are fungible quantities in this case. An analogy can be drawn between the classifiers used in [**?**] and the bandit arms used by Auer and Ortner[**?**].

Thompson and Wettergreen [**?**] maximize diversity of collected samples by using mutual information sampling. This approach ensures diversity in the collected sample set, an act that reduces uncertainty in the input space of a function. Neither mutual information nor maximum entropy sampling methods, when used with stationary Gaussian processes, take into account the dependent variable when selecting samples.

The prior work described above assumes one is choosing among a number of options and want to choose the maximally informative one. While choosing the maximally informative option is a useful guiding principle when robot explorers are presented with a number of sampling opportunities, it does not address the problem that explorers may have to give up a sampling opportunity in the hopes of finding better ones. Further it is not guaranteed that there is no cost associated with getting to sampling opportunities, an assumption commonly made when querying an oracle.

The prior work yields two observations. Firstly, foraging, a better model for planetary exploration, requires a measure of value of the sampling opportunities available to the ex-

ploring agent. Secondly, active learning uses uncertainty – in both input and output space of a function – to value potential exploration opportunities. What follows next is a method for exploring that reflects the limitations of a planetary setting and incorporates the result of sampling operations into decision making processes.

5. Method

The experiment builds on prior work.

- Combining foraging models with bandit literature - Previous work had a limit on the number of samples it could take - This experiment models a type of prospecting where the number of samples isn't limited but they do take time. - To that end we are looking at productivity.

- This experiment is more akin to contextual bandits. - The image represents a context, the NIRVSS - Apply texturecam classification of a scene, as the context - the choice is to sample or continue

- Productivity

6. Results

7. Conclusion

Appendix A. Section in Appendix

Sample text. Sample text. Sample text. Sample text. Sample text. Sample text. Sample text. Sample text. Sample text. Sample text. Sample text. Sample text. Sample text.