Near-optimal Observation Selection using Submodular Functions

Andreas Krause
Carnegie Mellon University

Carlos Guestrin
Carnegie Mellon University

Abstract

AI problems such as autonomous robotic exploration, automatic diagnosis and activity recognition have in common the need for choosing among a set of informative but possibly expensive observations. When monitoring spatial phenomena with sensor networks or mobile robots, for example, we need to decide which locations to observe in order to most effectively decrease the uncertainty, at minimum cost. These problems usually are NP-hard. Many observation selection objectives satisfy *submodularity*, an intuitive diminishing returns property – adding a sensor to a small deployment helps more than adding it to a large deployment. In this paper, we survey recent advances in systematically exploiting this submodularity property to *efficiently* achieve *near-optimal* observation selections, under complex constraints. We illustrate the effectiveness of our approaches on problems of monitoring environmental phenomena and water distribution networks.

Introduction

In many artificial intelligence applications, we need to effectively collect information in order to make best decisions under uncertainty. In this setting, we usually need to trade-off the informativeness of the observation and the cost of acquiring the information. When monitoring spatial phenomena using sensor networks, for example, we can decide where to place sensors. Since we have a limited budget, we want to place the sensors only at the most informative locations. Hence we want to select a set $\mathcal{A} \subseteq \mathcal{V}$ of locations, and want to maximize some objective $F(\mathcal{A})$ measuring the informativeness of the selected locations, subject to a constraint on the number of sensors we can place, i.e., $|\mathcal{A}| \leq k$. If we collect information using mobile robots, or if the placed sensors need to communicate wirelessly, we have more complex constraints on how we can make these observations. In the multi-robot case, for example, the chosen locations must lie on a collection of paths. In the wireless communication case, the locations need to be close enough to enable efficient wireless communication. These optimization problems are generally NP-hard (Krause & Guestrin 2005b). Therefore, heuristic approaches have commonly been applied, which cannot provide performance guarantees. In our recent work (Krause & Guestrin 2005c; Guestrin, Krause, & Singh 2005; Krause *et al.* 2006; 2007; Singh *et al.* 2007), we have presented several efficient algorithms for approximately solving these optimization problems. In contrast to the heuristic approaches, our algorithms have rigorous theoretical performance guarantees. In order to achieve these guarantees, we exploit a key property of many natural observation selection objectives: In most problems, adding an observation helps more, if we have made few observations so far, and helps less if we already have made many observations. This intuitive diminishing returns property is formalized by the concept of *submodularity*, which will be introduced in the following. We will illustrate the methodology on two important observation selection problems we have considered in the past, environmental monitoring, and securing water distribution networks.

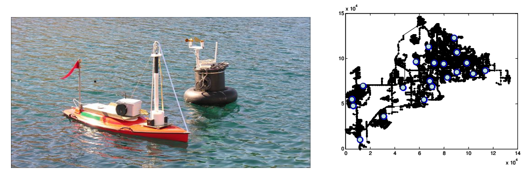


Figure 1: (a) Monitoring Lake Fulmor [courtesy of NAMOS (http://robotics.usc.edu/∼namos) & CENS (http://cens.ucla.edu)]; (b) Water network sensor placement in BWSN challenge.

Applications and Selection Objectives

Environmental Monitoring

Consider, for example, the monitoring of algae biomass in a lake. High levels of pollutants, such as nitrates, can lead to the development of algal blooms. These nuisance algal blooms impair the beneficial use of aquatic systems. Measuring quantities, such as pollutants, nutrients, and oxygen levels, can provide biologists with a fundamental characterization of the ecological state of such a lake. Unfortunately, such sensors are expensive, and it is impractical to cover the lake with these devices. Hence, a set of robotic boats (as in Fig. 1 (a)) have been used to move such sensors to various locations in the lake (Dhariwal *et al.* 2006). In order to make most effective measurements, we want to move the robots, such that the few observed values help us predict the algae biomass everywhere in the lake as well as possible.

Geometric objectives. A common approach for observation selection has been to use a geometric approach. With every potential sensing location $s \in \mathcal{V}$, one associates a geometric shape R_s, a *sensing region*, which is usually taken as a disk (*c.f.,* Bai *et al.* 2006), or a cone (for modeling camera viewing fields). Once an observation has been made, all points in the sensing region R_s are considered observed. The objective function is then $F_G(\mathcal{A}) = |\bigcup_{s\in\mathcal{A}} R_s|$, where $|\bigcup_{s\in\mathcal{A}} R_s|$ is the cardinality (or volume) of the covered set.

Probabilistic objectives. The assumption made by geometric objectives is that every location is either fully observed or unobserved. Often, observations are noisy, and by combining several observations we can achieve better coverage. To address this uncertainty, the spatial phenomena is often modeled probabilistically. In the above example, we would discretize the lake into finitely many locations $\mathcal{V}$, associate a random variable $\mathcal{X}_s$ with every location $s \in \mathcal{V}$ and describe a joint distribution $P(\mathcal{X}_\mathcal{V}) = P(\mathcal{X}_{s_1}, \dots, \mathcal{X}_{s_n})$ over all random variables. We can then use observations at a set $\mathcal{X}_\mathcal{A} = \mathbf{x}_\mathcal{A}$ to predict the phenomenon everywhere, by considering the conditional distributions $P(\mathcal{X}_{\mathcal{V}\setminus\mathcal{A}} \mid \mathcal{X}_\mathcal{A} = \mathbf{x}_\mathcal{A})$. We can also use this conditional distribution to quantify the uncertainty in the prediction. A good observation selection will have small uncertainty in the prediction everywhere.

Several such objective functions have been considered in the literature. In (Guestrin, Krause, & Singh 2005), we consider the *mutual information* between a set of chosen locations and their complement. This criterion is defined as $F_{\mathrm{MI}}(\mathcal{A}) = I(\mathcal{X}_\mathcal{A}; \mathcal{X}_{\mathcal{V}\setminus\mathcal{A}}) = H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{A}}) - H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{A}} \mid \mathcal{X}_\mathcal{A})$. It hence measures the decrease in Shannon entropy in the unobserved locations (prior uncertainty $H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{A}})$) achieved by making observations, leading to posterior uncertainty $H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{A}} \mid \mathcal{X}_\mathcal{A})$. As a model, we use Gaussian Processes (GPs), which have been frequently used to model spatial phenomena (*c.f.,* Cressie 1991). The distribution $P(\mathcal{X}_\mathcal{V})$ is then a multivariate normal distribution and the mutual information can be computed in closed form.

Securing Water Distribution Systems

Water distribution networks (as in Fig. 1 (b)), which bring the water to our taps, are complex dynamical systems, with impact on our everyday lives. Accidental or malicious introduction of contaminants in such networks can have severe impact on the population. Such intrusions could potentially be detected by a network of sensors placed in the water distribution system. The high cost of the sensors makes optimal placement an important issue. We recently participated in the *Battle of Water Sensor Networks (BWSN)*, an international challenge for designing sensor networks in several realistic settings. A set of *intrusion attacks* S was considered; each attack refers to the hypothetical introduction of contaminants at a particular node of the network, at a particular time of day, and for a specified duration. Using EPANET 2.0 (Rossman 1999), a simulator provided by the EPA, the impact of any given attack can be simulated, as well as, for any considered sensor placement $\mathcal{A} \subseteq \mathcal{V}$, the expected time to detection and estimated affected population computed. An optimal sensor placement minimizes these objectives.

Minimizing adverse effects. In the water distribution setting, we want to minimize adverse effects caused by a contaminant introduction. For every possible contamination attack $i \in S$, we can compute the *penalty* $\pi_i(t)$ incurred when detecting the intrusion at time t (where $t = 0$ at the start of the simulation). For example, π_i can measure the estimated population affected (across the entire network) by attack i at time t. For a sensor $s \in \mathcal{V}$ and attack $i \in S$, we can use the simulation to determine the time of detection, $T(s, i)$. Naturally, for a set of sensors, $T(\mathcal{A}, i) = \min_{s\in\mathcal{A}} T(s, i)$. We can then define the *penalty reduction* for a sensor placement $\mathcal{A} \subseteq \mathcal{V}$ as $R_i(\mathcal{A}) = \pi_i(T_{\max}) - \pi_i(T(\mathcal{A}, i))$. The final objective is then the *expected* penalty reduction, where the expectation is taken w.r.t. a probability distribution P over the attacks: $F_R(\mathcal{A}) = \sum_i P(i) R_i(\mathcal{A})$.

Submodularity of Observation Selection

In the previous section, we presented a collection of practical observation selection objectives, F_G, F_{MI}, and F_R. All these objectives (and many other practical ones) have the following key property in common: Adding an observation helps more if we have made few observations so far and helps less if we have made many observations. This diminishing returns property is visualized in Fig. 2(c). The score obtained is a concave function of the number of observations made. This effect is formalized by the concept of *submodularity* (*c.f.,* Nemhauser, Wolsey, & Fisher 1978). A real-valued function F, defined on subsets $\mathcal{A} \subseteq \mathcal{V}$ of a finite set $\mathcal{V}$ is called *submodular* if for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and for all $s \in \mathcal{V}\setminus\mathcal{B}$, it holds that $F(\mathcal{A}\cup\{s\}) - F(\mathcal{A}) \geq F(\mathcal{B}\cup\{s\}) - F(\mathcal{B})$. Hence, many observation selection problems can be reduced to the problem of maximizing a submodular set function subject to some constraints (e.g., the number of sensors we can place). Constrained maximization of submodular functions (and all objective functions discussed here) in general is NP-hard. The following sections survey efficient approximation algorithms with provable quality guarantees.

Optimizing Submodular Functions

Cardinality and Budget Constraints. We first consider the problem where each location $s \in \mathcal{V}$ has a fixed positive cost $c(s)$, and the cost of an observation selection $\mathcal{A} \subseteq \mathcal{V}$, $c(\mathcal{A})$ is defined as $c(\mathcal{A}) = \sum_{s\in\mathcal{A}} c(s)$. The problem then is to solve the following optimization problem:

$$\mathcal{A}^* = \operatorname{argmax}_\mathcal{A} F(\mathcal{A}) \text{ subject to } c(A) \leq B, \qquad (1)$$

for some nonnegative budget B. The special case $c(A) = |A|$ is called the *unit cost* case.

A natural heuristic which has been frequently used for the unit cost case is the *greedy algorithm*. This algorithm starts with the empty set $\mathcal{A} = \emptyset$, and iteratively, in round j, it finds the location $s_j = \operatorname{argmax}_s F(\mathcal{A} \cup \{s\}) - F(\mathcal{A})$, i.e., the location which increases the objective the most, and adds it to the current set $\mathcal{A}$. The algorithm stops after B sensors have been chosen. Surprisingly, this straight-forward heuristic has strong theoretical guarantees:

Theorem 1 (Nemhauser, Wolsey, & Fisher (1978)) *If F is a submodular, nondecreasing set function and $F(\emptyset) = 0$, then the greedy algorithm is guaranteed to find a set $\mathcal{A}$, such that $F(\mathcal{A}) \geq (1 - 1/e) \max_{|\mathcal{A}|=B} F(\mathcal{A})$.*

Hence, the greedy solution achieves an objective value that achieves at least a factor $(1 - 1/e)$ ($\approx 63\%$) of the optimal score. In order to apply Theorem 1, we must verify that the objective functions are indeed nondecreasing and $F(\emptyset) = 0$.

[1] F_{MI} is *approximately* nondecreasing (Guestrin, Krause, & Singh 2005), which preserves the approximation guarantee up to an arbitrarily small additive error.

A set function is called *nondecreasing* if for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ it holds that $F(A) \leq F(B)$. All objective functions discussed above are nondecreasing[1], and satisfy $F(\emptyset) = 0$.

The results extend to cases, where $c(s)$ is arbitrary. Here, a slight modification of the algorithm, combining the greedy selection of the location with highest benefit-cost ratio $s_j = \operatorname{argmax}_s \frac{F(\mathcal{A}\cup\{s\})-F(\mathcal{A})}{c(s)}$, with a partial enumeration scheme achieves the same constant factor $(1 - 1/e)$ approximation guarantee (Sviridenko 2004; Krause & Guestrin 2005a).

Path Constraints. In the robotic lake monitoring example described above, the observation selection problem is even more complex: Here, the locations $s \in \mathcal{V}$ are nodes in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, d)$, with edge weights $d : \mathcal{E} \to \mathbb{R}$ encoding distance. The goal is to find a collection of paths $\mathcal{P}_i$ in this graph, one for each of the k robots, providing highest information $F(\mathcal{P}_1 \cup \cdots \cup \mathcal{P}_k)$, subject to a constraint on the path length. Fig. 2(a) visualizes this setting. In this setting, the simple greedy algorithm – selecting most informative observations, such that the robot always keeps enough fuel to return to the goal – performs arbitrarily badly.

In (Singh *et al.* 2007), we show that the multiple robot problem can be reduced to optimizing paths for a single robot. We prove that if we have *any* single robot algorithm achieving an approximation guarantee of η (i.e., the solutions are at most a fraction $(1/\eta)$ from optimal), a simple sequential allocation strategy – iteratively optimizing one path at a time – achieves (nearly) the same approximation guarantee (η+1) for the case of multiple robots.

Chekuri & Pal (2005) proposed an algorithm for optimizing a single path $\mathcal{P}$ subject to a path-length constraint, achieving near-maximum submodular objective value $F(\mathcal{P})$. This algorithm achieves an approximation guarantee of $\eta = \log|\mathcal{P}^*|$, where $\mathcal{P}^*$ is the the optimal path. Unfortunately, the algorithm is only pseudopolynomial; its running time is $\mathcal{O}((|\mathcal{V}|B)^{\log|\mathcal{V}|})$. In (Singh *et al.* 2007), in addition to generalizing this algorithm to multiple robots with guarantee $\eta+1$, we present a spatial decomposition approach and devise branch and bound schemes, which make the algorithm of Chekuri & Pal practical. In our approach, the lake is partitioned into a grid of cells. These cells are considered nodes in a new graph with far fewer nodes, and the algorithm of Chekuri et.al. is applied on this new graph. Using an adapted version of our approach, computational cost can be traded off with the achieved objective value.

Theorem 2 *Let $\mathcal{P}^*$ be the optimal solution for the single robot problem with a budget of B and spatial decomposition into M cells of size L. Then our algorithm finds a solution $\widehat{\mathcal{P}}$ with information value of at least $F(\widehat{\mathcal{P}}) \geq \frac{1-1/e}{1+\log_2 N} F(\mathcal{P}^*)$, whose cost is no more than $\mathcal{O}(LB)$, in time $\mathcal{O}((MB)^{\log M})$.*

Communication Constraints. Often, when placing sensors, the sensors must be able to communicate with each other. In the case of wireless sensor networks (WSN), for example, the sensors need to communicate through lossy links which deteriorate with distance, causing message loss and increased power consumption. We can also model this problem by considering the potential locations as nodes in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, d)$. In (Krause *et al.* 2006), we assign the *expected number of retransmissions* (i.e., the expected number of times a message has to be resent in order to warrant successful transmission) between locations u and v as edge cost $d(u, v)$. The cost of a set of locations $\mathcal{A} \subset \mathcal{V}$ is then the minimum cost of connecting the locations $\mathcal{A}$ in the graph $\mathcal{G}$. More precisely, the cost $c(\mathcal{A})$ of a placement is the sum of the edge costs of the cheapest (Steiner) tree containing the nodes in $\mathcal{A}$. Under this new cost function, we consider the maximization problem (1). Here, as in the path constraint case, the greedy algorithm performs arbitrarily poorly. In (Krause *et al.* 2006), we present a randomized approximation algorithm for approximately solving this problem. This algorithm exploit an additional property of many observation selection objectives: Intuitively, sensors which are far apart make roughly independent observations. Formally, require that there are constants $r \geq 0$ and $0 < \gamma \leq 1$, such that, if two sets of locations $\mathcal{A}$ and $\mathcal{B}$ are at least distance r apart, it holds that $F(\mathcal{A}\cup\mathcal{B}) \geq F(\mathcal{A}) + \gamma F(\mathcal{B})$. In this case, we say that F is (r, γ)-local. This *locality* property is empirically satisfied for mutual information F_{MI} as we show in (Krause *et al.* 2006). For geometric objectives, F_G is (r, γ) local for $\gamma = 1$ and $r = 2\max_s \operatorname{diam}(R_s)$.

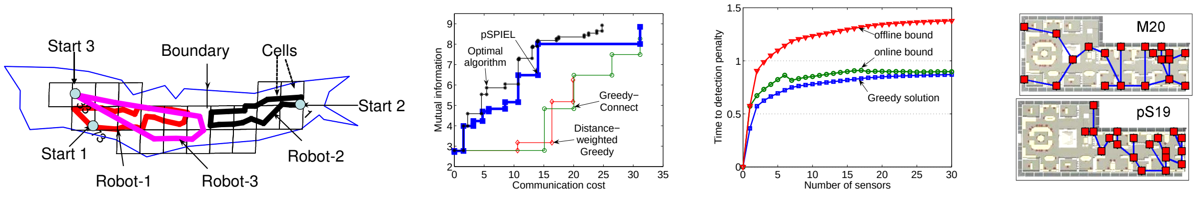
Theorem 3 *Given a graph $\mathcal{G} = (\mathcal{V}, E, d)$, and an (r, γ)-local monotone submodular function F, we can find a tree $\mathcal{T}$ with cost $\mathcal{O}(r\log|\mathcal{V}|) \times c(\mathcal{T}^*)$, spanning a set $\mathcal{A}$ with $F(\mathcal{A}) \geq \Omega(\gamma) \times F(\mathcal{A}^*)$. The algorithm is randomized and runs in polynomial-time.*

Theorem 3 shows that we can solve the problem (1) to provide a sensor placement for which the communication cost is at most a small factor (at worst logarithmically) larger, and for which the expected sensing quality is at most a constant factor worse than the optimal solution. Fig. 2(b) compares the performance of our approximation algorithm – pSPIEL – with the (intractable) optimal algorithm (exhaustive search), as well as two reasonable heuristics, when optimizing a WSN to monitor temperature in a building. pSPIEL performs much closer to the optimal solution.

We used pSPIEL to design a WSN for monitoring light in the Intelligent Workplace at CMU. Fig. 2(d) compares a manual deployment of 20 sensor motes with 19 motes deployed by pSPIEL. Surprisingly, pSPIEL does not extend the placement in the Western area of the building. During data collection, the Western part was unused at night; hence accurate indoor light prediction requires many sensors in the Eastern part. Daylight intensity can be well predicted even without sensors in the Western part. pSPIEL lead to a 30% decrease in RMS error, and reduced the communication cost.

Extensions

Online guarantees for observation selection. The $(1 - 1/e)$ bound for the greedy algorithm is *offline*, as we can state it before running the algorithm. However, we can exploit submodularity even further to compute – often much tighter – *online* bounds on the optimal solution. We look at the current solution $\mathcal{A}$, obtained, for example, by the greedy or any other algorithm. For each observation s which has not been considered yet, let $\delta_s = F(\mathcal{A}\cup\{s\}) - F(\mathcal{A})$ be the improvement in score we would get by adding observation



(a) *Path constraints* (b) *Communic. constraints* (c) *Submodular returns* (d) *Placements*

Figure 2: (a) Informative paths planned for 3 boats on Lake Fulmor. (b) pSPIEL achieves near-optimal cost-benefit ratio. (c) Submodularity gives tight online bounds for any selection. (d) WSN deployments in CMU's Intelligent Workplace (M20: manual, pS19: using pSPIEL).

s. Assume we want to select k observations. Let $s_1, \dots, s_k$ be the k observations for which δ_s is largest. Then $F(\mathcal{A}^*) \le F(\mathcal{A}) + \sum_{j=1}^{k} \delta_{s_j}$. This bound allows us to get guarantees about *arbitrary* observation selections $\mathcal{A}$. Fig. 2(c) shows offline and online bounds achieved for placing an increasing number of sensors, when optimizing the time to detection when securing water networks. The online bound is much tighter, and closer to the score achieved by the greedy algorithm. These bounds also allows us to use mixed-integer programming (MIP) (Nemhauser & Wolsey 1981), which can be used to solve for the optimal observation selection, or to acquire even tighter bounds on the optimal solution by computing the linear programming relaxation.

Fast implementation. Submodularity can also be used to speed up the greedy algorithm. The key observation is that the incremental benefits $\delta_s(\mathcal{A}) = F(\mathcal{A} \cup \{s\}) - F(\mathcal{A})$ are monotonically nonincreasing in $\mathcal{A}$. By exploiting this observation, we can "lazily" recompute only the $\delta_s(\mathcal{A})$ only for the "most promising" locations s (Robertazzi & Schwartz 1989). On the large BWSN network, this lazy implementation decreased the running time from 30 hours to 1 hour.

Robust Sensor Placement. Sensor nodes are susceptible to failures, e.g., through loss of power. In (Krause *et al.* 2007), we describe an approach for optimizing placements which are robust against failures. The key observation is that submodular functions are closed under nonnegative linear combinations. Hence, the expected placement score over all possible failure scenarios is submodular. Similarly, this observation can also be used to handle uncertainty in the model (network links, parameter uncertainty, etc.).

Multicriterion Optimization. Often, we have multiple objectives $F_1, \dots, F_m$ which we want to simultaneously optimize. In the water network example, we want to simultaneously maximize the likelihood of detecting an intrusion, as well as minimize the expected population affected by an intrusion. In general, two placements $\mathcal{A}$ and $\mathcal{B}$ might be *incomparable*, i.e., neither $\mathcal{A}$ nor $\mathcal{B}$ is better in all m criteria. Hence, the goal in multicriterion optimization is to find *Pareto-optimal* placements $\mathcal{A}$, i.e., those such that there cannot exist a placement $\mathcal{B}$ for which $F_i(\mathcal{B}) \ge F_i(\mathcal{A})$ for $1 \le i \le m$, and $F_j(\mathcal{B}) > F_j(\mathcal{A})$ for some j. A common technique for finding such Pareto-optimal solutions is *scalarization*, i.e., choosing a set of positive weights $\lambda_1, \dots, \lambda_m > 0$ and maximizing $F(\mathcal{A}) = \sum_i \lambda_i F_i(\mathcal{A})$. Since positive linear combinations of submodular functions are still submodular, the resulting scalarized problem is a submodular maximization problem, which can be addressed using algorithms surveyed in this paper.

Implications for AI

Observation selection problems are ubiquitous in AI. We now discuss several examples of fundamental AI problems which could potentially be addressed using our algorithms.

Fault Diagnosis. A classical AI problem is probabilistic fault diagnosis as considered, e.g., by Zheng, Rish, & Beygelzimer (2005). Here, one wants to select a set of tests to probe the system (e.g., a computer network), in order to diagnose the state $\mathcal{X}_\mathcal{U}$ of unobservable system attributes $\mathcal{U}$ (e.g., presence of a fault). As criterion of informativeness, Zheng, Rish, & Beygelzimer (2005) use the information gain $F_{\mathrm{IG}}(\mathcal{A}) = I(\mathcal{X}_\mathcal{A}; \mathcal{X}_U)$ of the selected tests $\mathcal{A} \subseteq \mathcal{V}$ about the target attributes $\mathcal{U}$. Krause & Guestrin (2005c) show that for a wide class of probabilistic graphical models, this criterion F_{IG} is submodular. Often, the cost of a test depends on the tests already chosen (i.e., test A might be cheaper if test B has been chosen). Algorithms as surveyed in this paper can potentially be used to approximately solve problems with such complex constraints. Related applications are test selection for expert systems, e.g., in the medical domain, as well as optimizing test cases for software testing coverage.

Robotic Exploration. An important problem in robotics is Simultaneous Localization and Mapping (SLAM) (*c.f.,* Sim & Roy (2005)). There, robots make observations in order to simultaneously localize themselves, and detect landmarks to create a map. Commonly, probabilistic models such as Kalman Filters and extensions thereof are used to track the uncertainty about the state variables $\mathcal{X}_\mathcal{U}$ (landmark and robot locations). Using such a probabilistic model, observations $\mathcal{A}$ can be actively collected by controlling the robot, with the goal of maximizing the information gain $I(\mathcal{X}_\mathcal{A}; \mathcal{X}_\mathcal{U})$ about the unobserved locations, as considered by Sim & Roy (2005). Since the robot's movement is constrained by obstacles, algorithms as those surveyed in this paper can potentially be used to plan such informative paths.

Minimizing Human Attention. In this paper, we considered problems where each observation is expensive to acquire. In many applications however, we have an abundance of information at our disposal, and the scarce resource is the attention of the human, to which this information should be presented. One example is active learning (*c.f.,* Hoi *et al.* (2006)), where the goal is to learn a classifier, but the training data is initially unlabeled. A human expert can be requested to label a set of examples, but each label is expensive. In some applications, complex constraints are present; e.g., when labeling a stream of video images, labeling a sequence of consecutive images is cheaper than labeling individual images. Hoi *et al.* (2006) show that certain active learning objectives are (approximately) submodular, hence our algorithms could potentially be used in this context.

Another example is information presentation. Here, we want to, e.g., choose a subset of email requests to display, or automatically summarize a large document, etc. In these problems, each selection of emails or sentences achieves a certain utility to the user, which is to be maximized. We also have complex constraints (e.g., reading long emails is more "expensive"; when reading an email in a communication thread, the initiating email has to be presented as well, etc.), which could potentially be addressed using our algorithms. Here, the key open research challenge is to understand, which classes of utility functions are submodular.

Influence Maximization. Kempe, Kleinberg, & Tardos (2003) show that the problem of selecting a set of people in a social network with maximum *influence* is a submodular optimization problem. Here, an initial set of people are, e.g., targeted by a marketing campaign. The probability that any person (initially untargeted) in the network becomes influenced, depends on how many of their neighbors are already influenced, as well as the strength of their interaction. Rather than simply asking for the best set of k people to target in order to maximize the influence on the whole network, one could use our algorithms to consider more complex constraints. For example, a marketing tour (e.g., an author signing books) could proceed through the country, and the cost to influence the next person depends on the tour's current location, and the problem would be to find the optimum such tour (or multiple tours at the same time).

Monotonic Constraint Satisfaction. A key problem in AI is finding satisfying assignments to logical formulas, or maximizing the number of satisfied clauses. Consider the special case, where we are given a boolean formula $f(x_1, \ldots, x_n) = C_1 \wedge \cdots \wedge C_m$, where each C_i is a clause with weight $w_i \geq 0$, and containing the disjunction of any set of positive literals (i.e., the formula does not contain negations). For a subset $\mathcal{A} \subseteq \{1, \ldots, n\}$ of indices of variables x_i set to *true*, the function $F(\mathcal{A}) = \sum_{i:C_i \text{ satisfied by some } x_i \in \mathcal{A}} w_i$ is monotonic and submodular. Hence, the problem of selecting a set $\mathcal{A}$ of k variables to set to *true* maximizing the weight of satisfied clauses is a submodular maximization problem. Our algorithms could also allow us to find such assignments subject to more complex constraints among the variables, e.g., where setting x_1 to true reduces the cost of setting x_4 to true.

Conclusions

We have reviewed recent work on optimizing observation selections. Many natural observation selection objectives are submodular. We demonstrated how submodularity can be exploited to develop efficient approximation algorithms with provable quality guarantees for observation selection. These algorithms can handle a variety of complex combinatorial constraints, e.g., requiring that the selected observations lie on a collection of paths. Submodularity also provides a posteriori bounds for any algorithm, and can be extended to, e.g., handle failing sensors, and multi-criterion optimization.

We are convinced that even beyond the problems considered in this paper, submodular optimization can help to tackle important AI search problems, and submodularity is a concept which can be more widely considered in AI.

References

Bai, X.; Kumar, S.; Yun, Z.; Xuan, D.; and Lai, T. H. 2006. Deploying wireless sensors to achieve both coverage and connectivity. In *MobiHoc*.

Chekuri, C., and Pal, M. 2005. A recursive greedy algorithm for walks in directed graphs. In *FOCS*, 245–253.

Cressie, N. A. C. 1991. *Statistics for Spatial Data*. Wiley.

Dhariwal, A.; Zhang, B.; Stauffer, B.; Oberg, C.; Sukhatme, G. S.; Caron, D. A.; and Requicha, A. A. 2006. Networked aquatic microbial observing system. In *ICRA*.

Guestrin, C.; Krause, A.; and Singh, A. 2005. Near-optimal sensor placements in Gaussian processes. In *ICML*.

Hoi, S. C. H.; Jin, R.; Zhu, J.; and Lyu, M. R. 2006. Batch mode active learning and its application to medical image classification. In *ICML*.

Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing the spread of influence through a social network. In *KDD*.

Krause, A., and Guestrin, C. 2005a. A note on the budgeted maximization of submodular functions. Technical report, CMU-CALD-05-103.

Krause, A., and Guestrin, C. 2005b. Optimal nonmyopic value of information in graphical models - efficient algorithms and theoretical limits. In *Proc. of IJCAI*.

Krause, A., and Guestrin, C. 2005c. Near-optimal value of information in graphical models. In *UAI*.

Krause, A.; Guestrin, C.; Gupta, A.; and Kleinberg, J. 2006. Near-optimal sensor placements: Maximizing information while minimizing communication cost. In *IPSN*.

Krause, A.; Leskovec, J.; Guestrin, C.; VanBriesen, J.; and Faloutsos, C. 2007. Efficient sensor placement optimization for securing large water distribution networks. *Submitted to the Journal of Water Resources Planning an Management*.

Nemhauser, G. L., and Wolsey, L. A. 1981. *Studies on Graphs and Discrete Programming*. North-Holland. chapter Maximizing Submodular Set Functions: Formulations and Analysis of Algorithms, 279–301.

Nemhauser, G.; Wolsey, L.; and Fisher, M. 1978. An analysis of the approximations for maximizing submodular set functions. *Mathematical Programming* 14:265–294.

Robertazzi, T. G., and Schwartz, S. C. 1989. An accelerated sequential algorithm for producing D-optimal designs. *SIAM Journal of Scientific and Statistical Computing* 10(2):341–358.

Rossman, L. A. 1999. The epanet programmer's toolkit for analysis of water distribution systems. In *Annual Water Resources Planning and Management Conference*.

Sim, R., and Roy, N. 2005. Global a-optimal robot exploration in slam. In *ICRA*.

Singh, A.; Krause, A.; Guestrin, C.; Kaiser, W.; and Batalin, M. 2007. Efficient planning of informative paths for multiple robots. In *IJCAI*.

Sviridenko, M. 2004. A note on maximizing a submodular set function subject to knapsack constraint. *Ops Res Lett* 32:41–43.

Zheng, A. X.; Rish, I.; and Beygelzimer, A. 2005. Efficient test selection in active diagnosis via entropy approximation. In *UAI*.