Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Submodular Surrogates for Value of Information

Yuxin Chen
ETH Zürich
yuxin.chen@inf.ethz.ch

Shervin Javdani
Carnegie Mellon University
sjavdani@cmu.edu

Amin Karbasi
Yale University
amin.karbasi@yale.edu

J. Andrew Bagnell
Carnegie Mellon University
dbagnell@ri.cmu.edu

Siddhartha Srinivasa
Carnegie Mellon University
ss5@andrew.cmu.edu

Andreas Krause
ETH Zürich
krausea@ethz.ch

Abstract

How should we gather information to make effective decisions? A classical answer to this fundamental problem is given by the decision-theoretic value of information. Unfortunately, optimizing this objective is intractable, and myopic (greedy) approximations are known to perform poorly. In this paper, we introduce DIRECT, an efficient yet near-optimal algorithm for nonmyopically optimizing value of information. Crucially, DIRECT uses a novel surrogate objective that is: (1) aligned with the value of information problem (2) efficient to evaluate and (3) adaptive submodular. This latter property enables us to utilize an efficient greedy optimization while providing strong approximation guarantees. We demonstrate the utility of our approach on four diverse case-studies: touch-based robotic localization, comparison-based preference learning, wild-life conservation management, and preference elicitation in behavioral economics. In the first application, we demonstrate DIRECT in closed-loop on an actual robotic platform.

Introduction

In many real-world decision making tasks we must adaptively choose among informative but expensive tests. As an illustrative example, consider medical diagnosis (Kononenko 2001), where many medical tests are available, and we aim to administer tests that will enable us to provide effective treatment. In such systems, the reward of making a decision depends on some unknown hidden state (e.g., the patient's condition). Generally, it is impossible to observe this hidden state directly, but one can perform tests, and observe the outcome of variables correlated with the hidden state, at some cost. The task is then to find a policy for selecting the most informative tests, so that we can gather enough information to make effective decisions, while minimizing the cost of testing. Similar problems arise in numerous other domains, ranging from optimal experimental design (Chaloner and Verdinelli 1995) to recommender systems (Javdani et al. 2014) to policy making (Runge, Converse, and Lyons 2011).

Related work A classical approach to information gathering for decision making is the decision-theoretic *value of information* (Howard 1966). Here, we seek policies that maximize the increase in the maximum expected utility that the decision maker could obtain when acting upon the acquired information. Optimizing this criterion in general probabilistic models is NP^{PP}-complete (Krause and Guestrin 2009). Recently, *Same-Decision Probability* (SDP) has been proposed for the purpose of robust decision making (Choi, Xue, and Darwiche 2012). However, it has been shown that SDP is PP^{PP}-complete in general, and even remains NP-hard in Navie Bayes Nets (Chen, Choi, and Darwiche 2014). Consequently, greedy heuristics that myopically select the next test are employed. It is known (Golovin, Krause, and Ray 2010) that these heuristics can perform arbitrarily poorly; unfortunately exact algorithms for *non-myopic* value of information have so far been restricted to simple probabilistic models (Krause and Guestrin 2009).

One can model the non-myopic value of information problem as a Partially Observable Markov Decision Process (POMDP) (Smallwood and Sondik 1973; Kaelbling, Littman, and Cassandra 1998), where state represents the selected tests and observed outcome of each test. Unfortunately, this gives us an exponentially large state space, making the application of many black-box POMDP solvers (e.g., (Pineau, Gordon, and Thrun 2006)) infeasible.

The problem of selecting information gathering tests for purely *reducing uncertainty about some hidden variable* (ignoring utilities of decision making) is studied in the context of active learning (Dasgupta 2004; Balcan, Beygelzimer, and Langford 2006; Hanneke 2007; Settles 2012) and (Bayesian) experimental design (Chaloner and Verdinelli 1995). Deriving optimal policies is generally NP-hard (Chakaravarthy et al. 2007), but some approximation results are known. In particular, if tests are noise-free (i.e., deterministic functions of the hidden state), the problem is known as the Optimal Decision Tree (ODT) problem, and a simple greedy algorithm, called generalized binary search (GBS), is guaranteed to produce a bounded approximation to the optimal policy in terms of the cost (Kosaraju, Przytycka, and Borgstrom 1999).

Recently, these results have been brought closer to decision making by associating each hidden state with some optimal decision(s). Information gathering policies no longer aim to reduce all uncertainty – but just enough to make the right decision. Two algorithms, namely *equiv-*

alence class edge cutting (EC^2) (Golovin, Krause, and Ray 2010) and *hyperedge cutting* (HEC) (Javdani et al. 2014) provide approximation guarantees for this problem. Since our approach builds on these techniques, we review them in more detail in the next section.

Our contributions In this paper, we provide a principled framework for a class of *non-myopic value of information* problems: We seek a minimum-cost policy which guarantees that, upon termination, a near-optimal decision – one that provides almost as much utility as achievable by carrying out *all* tests – is identified. Instead of optimizing for this directly, we construct DIRECT, a surrogate objective function with a few key properties. Crucially, we show that it exhibits *adaptive submodularity* (Golovin and Krause 2011), a natural diminishing returns property, generalizing the classical notion of submodularity to adaptive policies. This result allows us to greedily maximize the surrogate, while still providing a strong theoretical guarantee. We evaluate our algorithm on four applications: touch-based localization with a robotic arm (Javdani et al. 2013), comparison-based preference learning (Karbasi, Ioannidis, and Massoulié 2011; 2012), adaptive management for biodiversity conservation (Runge, Converse, and Lyons 2011), and preference elicitation in behavioral economics (Ray et al. 2012). Experimental results show that our algorithm significantly outperforms myopic value of information in most settings. Moreover, our algorithm is exponentially faster than HEC in theory, significantly faster (often by orders of magnitude) in practice, while offering similar empirical performance.

Background and Problem Statement

We now formalize the problem addressed in this paper – efficient information gathering for decision making – and review existing approaches for solving it.

The Value of Information and Decision Region Determination Problem

Assume that there is some unknown hidden discrete random variable $Y \in \mathcal{Y}$ upon which we want to make a decision. In our medical diagnostics example, Y may represent the condition of the patient. We are given a set $\mathcal{T} = \{1, \dots, n\}$ of possible (e.g., medical) tests; performing each test $t \in \mathcal{T}$ incurs a certain cost of $c(t) > 0$ and produces an outcome that is correlated with Y. We model the outcome of each test t by a discrete random variable $X_t \in \mathcal{X}$ and denote its observed outcome by x_t. Hereby, $\mathbf{x}_\mathcal{A} \in \mathcal{X}^\mathcal{A}$ is a vector of outcomes indexed by a set of tests $\mathcal{A} \subseteq \mathcal{T}$ that we have performed, and y is the realized value of the hidden variable Y. Further assume that there is a known prior distribution $\mathbb{P}[Y, X_1, \dots, X_n]$ over the hidden variable and test outcomes admitting efficient inference, i.e., we can compute the posterior distribution $\mathbb{P}[Y = y \mid \mathbf{x}_\mathcal{A}]$ efficiently after having observed any $\mathbf{x}_\mathcal{A}$.

Suppose there is a finite set $\mathcal{D}$ of decisions to choose from. After performing a set of tests and observing their outcomes, we want to make the best decision given our belief about the hidden variable Y (e.g., we must decide how to treat the patient). Formally, we quantify the benefit of making a decision $d \in \mathcal{D}$ for any $y \in \mathcal{Y}$ by a utility function $u : \mathcal{Y} \times \mathcal{D} \to \mathbb{R}_{\geq 0}$. The expected value of a decision d after observing $\mathbf{x}_\mathcal{A}$ is $U(d \mid \mathbf{x}_\mathcal{A}) = \mathbb{E}_y[u(y, d) \mid \mathbf{x}_\mathcal{A}]$. The value of a specific set of observations $\mathbf{x}_\mathcal{A}$ is then defined as: $\mathrm{VoI}(\mathbf{x}_\mathcal{A}) = \max_{d \in \mathcal{D}} U(d \mid \mathbf{x}_\mathcal{A})$, i.e., the maximum expected utility achievable when acting upon observations $\mathbf{x}_\mathcal{A}$.

Consider performing *all* tests, receiving outcomes $\mathbf{x}_\mathcal{T}$, and making the most informed decision possible. This would achieve a value of $\mathrm{VoI}(\mathbf{x}_\mathcal{T})$. However, it may be possible to achieve nearly $\mathrm{VoI}(\mathbf{x}_\mathcal{T})$ with far fewer tests. Our goal is to adaptively select the cheapest tests to do so. Formally, we define the *regret*[1] of a decision d given observations $\mathbf{x}_\mathcal{A}$ by $R(d \mid \mathbf{x}_\mathcal{A}) = \max_{\mathbf{x}_\mathcal{T} : \mathbb{P}[\mathbf{x}_\mathcal{T} \mid \mathbf{x}_\mathcal{A}] > 0} [\mathrm{VoI}(\mathbf{x}_\mathcal{T}) - U(d \mid \mathbf{x}_\mathcal{T})]$. This regret bounds our loss in expected utility if we stop upon observing $\mathbf{x}_\mathcal{A}$ and committing to action d. Our goal is to find a policy π of minimum cost with regret of at most ε. Formally, a policy is a partial mapping from observation vectors $\mathbf{x}_\mathcal{A}$ to tests, specifying which test to run next (or that we should stop testing if $\mathbf{x}_\mathcal{A}$ is not in the domain of π) for any observation vector $\mathbf{x}_\mathcal{A}$. If variables $X_1, \dots, X_n$ would result in outcomes $\mathbf{x}_\mathcal{T}$, we will obtain a set of observations, denoted as $\mathcal{S}(\pi, \mathbf{x}_\mathcal{T}) \subseteq \mathcal{T} \times \mathcal{X}$, by running policy π until termination (likely before exhausting all tests). The expected cost of a policy π is $\mathrm{cost}(\pi) = \mathbb{E}_{\mathbf{x}_\mathcal{T}}[c(\mathcal{S}(\pi, \mathbf{x}_\mathcal{T}))]$, where $c(\mathcal{S}(\pi, \mathbf{x}_\mathcal{T}))$ is the total cost of all tests run by π in the event $\mathbf{x}_\mathcal{T}$. Fix some small tolerance $\varepsilon \geq 0$. We seek a policy π^* with minimum cost, such that upon termination, π^* will suffer regret of at most ε:

$$\pi^* \in \arg\min_\pi \mathrm{cost}(\pi), \text{ s.t.}$$
$$\forall \mathbf{x}_\mathcal{T}\ \exists d : R(d \mid \mathcal{S}(\pi, \mathbf{x}_\mathcal{T})) \leq \varepsilon \text{ whenever } \mathbb{P}[\mathbf{x}_\mathcal{T}] > 0. \quad (1)$$

In other words, we require that each feasible policy satisfies the following condition: Upon termination, we must be able to commit to a decision, such that we lose *at most* ε expected utility, compared to the optimal decision we could have made if we had also observed *all remaining* unobserved variables. We call Problem (1) the *nonmyopic value of information problem for achieving near-maximal utility (NVOI-NMU)*.[2]

Importantly, this problem reduces[3] to a problem known as the *Decision Region Determination* (DRD) problem (Javdani et al. 2014). In DRD, we are given (1) a set of hypotheses $\mathcal{H} = \{h_1, \dots, h_N\}$; (2) a random variable H distributed over $\mathcal{H}$ with known distribution $\mathbb{P}$; (3) a set of tests modeled as deterministic functions $f_1, \dots f_n : \mathcal{H} \to \mathcal{X}$; (4) a cost function $c : \{1, \dots, n\} \to \mathbb{R}_+$ and (5) a collection of subsets $\mathcal{R}_1, \dots, \mathcal{R}_m \subseteq \mathcal{H}$ called *decision regions*. We seek a policy π^* of minimum cost, which adaptively picks tests i, observes their outcomes $X_i = f_i(H)$, where $H \in \mathcal{H}$ is the unknown hypothesis, such that upon termination, there exists at least

[1] Clearly, this regret is also an upper bound on the *expected* loss in expected utility, i.e., $\mathbb{E}_{\mathbf{x}_\mathcal{T}}[\mathrm{VoI}(\mathbf{x}_\mathcal{T}) - U(d \mid \mathbf{x}_\mathcal{T}) \mid \mathbf{x}_\mathcal{A}]$.

[2] In classical value of information, costs and utilities have the same units, and we aim to maximize benefit minus cost. In many cases (e.g. medical diagnosis), this is not the case, so we formulate our problem to achieve near-maximal utility with minimum cost.

[3] The NVOI-NMU and DRD problems are in fact equivalent.

one decision region that contains all hypotheses consistent with the observations made by the policy. That is, we seek

$$\pi^* \in \arg\min_{\pi} \text{cost}(\pi), \text{s.t. } \forall h \; \exists d : \mathcal{H}(\mathcal{S}(\pi, h)) \subseteq \mathcal{R}_d. \quad (2)$$

Hereby $h \in \mathcal{H}$, and $\mathcal{H}(\mathbf{x}_\mathcal{A}) = \{h' \in \mathcal{H} : (i, x) \in \mathbf{x}_\mathcal{A} \Rightarrow f_i(h') = x\}$ is the set of hypotheses consistent with $\mathbf{x}_\mathcal{A}$. To reduce the NVOI-NMU Problem (1) to DRD (2), we interpret every outcome vector $\mathbf{x}_\mathcal{T}$ with positive probability as a hypothesis h. The interpretation of the prior, tests, and costs follow immediately. It remains to define the decision regions. For each decision d, we set $\mathcal{R}_d$ to be the set of outcome vectors, for which d is an ε-optimal action, or formally: $\mathcal{R}_d = \{\mathbf{x}_\mathcal{T} : U(d \mid \mathbf{x}_\mathcal{T}) \geq \text{VoI}(\mathbf{x}_\mathcal{T}) - \varepsilon\}$.

Existing approaches for solving the DRD problem

As a special case of the Decision Region Determination problem, the *Equivalence Class Determination* (ECD) problem (Golovin, Krause, and Ray 2010) only allows *disjoint* decision regions, i.e., $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$ for $i \neq j$. This means that each hypothesis h is associated with a unique decision. The EC^2 algorithm (Golovin, Krause, and Ray 2010) considers hypotheses as nodes in a graph $G = (V, E)$, and defines weighted edges between hypotheses in different decision regions: $E = \cup_{i \neq j} \{\{h, h'\} : h \in \mathcal{R}_i, h' \in \mathcal{R}_j\}$, where the weight of an edge is defined as $w(\{h, h'\}) = \mathbb{P}[h] \cdot \mathbb{P}[h']$; similarly, the weight of a set of edges is $w(E') = \sum_{e \in E'} w(e)$. An edge is consistent with the observation iff both hypotheses incident to the edge are consistent. Hence, a test t with outcome x_t is said to cut edges $E(x_t) = \{\{h, h'\} \in E : f_t(h) \neq x_t \vee f_t(h') \neq x_t\}$. Performing tests will cut edges inconsistent with the observed test outcomes, and we aim to eliminate all inconsistent edges while minimizing the expected cost incurred.

The EC^2 objective is defined as the total weight of edges cut: $f_{EC}(\mathbf{x}_\mathcal{A}) := w\Big(\bigcup_{t \in \mathcal{A}} E(x_t)\Big)$. Let $\mathbb{P}[\mathcal{R}_i]$ be the total prior probability mass of all hypotheses h in $\mathcal{R}_i$. Then the weight of edges between distinct decision regions $\mathcal{R}_i, \mathcal{R}_j$ is $w(\mathcal{R}_i \times \mathcal{R}_j) = \sum_{h \in \mathcal{R}_i, h' \in \mathcal{R}_j} \mathbb{P}[h]\,\mathbb{P}[h'] = \mathbb{P}[\mathcal{R}_i]\,\mathbb{P}[\mathcal{R}_j]$. Naively, computing the total edge weight requires enumerating all pairs of regions. However, we can compute this in *linear* time by noting it is equivalent to an *elementary symmetric polynomial* of degree 2: $\sum_{i \neq j} w(\mathcal{R}_i \times \mathcal{R}_j) = \frac{1}{2}\left(\left(\sum_i \mathbb{P}[\mathcal{R}_i]\right)^2 - \sum_i \mathbb{P}[\mathcal{R}_i]^2\right)$. We similarly compute the total edge weight after observations $\mathbf{x}_\mathcal{A}$ using $\mathbb{P}[\mathcal{R}_i \cap \mathcal{H}(\mathbf{x}_\mathcal{A})]$ for the probability mass of all hypotheses in $\mathcal{R}_i$ consistent with observations $\mathbf{x}_\mathcal{A}$. Finally, we subtract these two quantities to compute $f_{EC}(\mathbf{x}_\mathcal{A}) = \sum_{i \neq j} w(\mathcal{R}_i \times \mathcal{R}_j) - \sum_{i \neq j} w(\mathcal{R}_i \cap \mathcal{H}(\mathbf{x}_\mathcal{A}) \times \mathcal{R}_j \cap \mathcal{H}(\mathbf{x}_\mathcal{A}))$.

EC^2 is known to be near-optimal for the ECD problem. This result relies on the fact that f_{EC} is *adaptive submodular*, and *strongly adaptive monotone* (Golovin and Krause 2011). Let $\mathbf{x}_\mathcal{A}$ and $\mathbf{x}_\mathcal{B}$ be two observation vectors. We call $\mathbf{x}_\mathcal{A}$ a *subrealization* of $\mathbf{x}_\mathcal{B}$, denoted as $\mathbf{x}_\mathcal{A} \preceq \mathbf{x}_\mathcal{B}$, if the index set $\mathcal{A} \subseteq \mathcal{B}$ and $\mathbb{P}[\mathbf{x}_\mathcal{B} \mid \mathbf{x}_\mathcal{A}] > 0$. A function $f : 2^{\mathcal{T} \times \mathcal{X}} \to \mathbb{R}$ is called *adaptive submodular* w.r.t. a distribution $\mathbb{P}$, if for any $\mathbf{x}_\mathcal{A} \preceq \mathbf{x}_\mathcal{B}$ and any test t it holds that $\Delta(t \mid \mathbf{x}_\mathcal{A}) \geq \Delta(t \mid \mathbf{x}_\mathcal{B})$, where $\Delta(t \mid \mathbf{x}_\mathcal{A}) := \mathbb{E}_{x_t}\big[f(\mathbf{x}_{\mathcal{A} \cup \{t\}}) - f(\mathbf{x}_\mathcal{A}) \mid \mathbf{x}_\mathcal{A}\big]$ (i.e., "adding information earlier helps more"). Further, function f is called *strongly adaptively monotone* w.r.t. $\mathbb{P}$, if for all $\mathcal{A}$, $t \notin \mathcal{A}$, and $x_t \in \mathcal{X}$, it holds that $f(\mathbf{x}_\mathcal{A}) \leq f(\mathbf{x}_{\mathcal{A} \cup \{t\}})$ (i.e., "adding information never hurts"). For decision problems satisfying adaptive submodularity and strongly adaptive monotonicity, the policy that greedily, upon having observed $\mathbf{x}_\mathcal{A}$, selects the test $t^* \in \arg\max_t \Delta(t \mid \mathbf{x}_\mathcal{A})/c(t)$, is guaranteed to attain near-minimal cost (Golovin and Krause 2011).

EC^2 crucially relies on the fact that decision regions are *disjoint*. In the presence of overlapping regions, there is no principled way to apply EC^2. Recently, the HEC algorithm (Javdani et al. 2014) was proposed for solving the general DRD problem. It does so by creating an alternate representation – a hypergraph for splitting decision regions. The computational bottleneck for HEC lies in the construction of this hypergraph, where computation cost grows *exponentially* with the hyperedge cardinality, which depends on the maximum number of optimal decisions one can make for a hypothesis. Thus, when we have large overlap between regions – the common case for NVOI-NMU, in particular with larger ε – HEC becomes intractable.

The Decision Region Edge Cutting Algorithm

We now develop an *efficient yet near-optimal* criterion, namely *Decision Region Edge Cutting* (DIRECT), for solving the DRD – and hence the NVOI-NMU – problem.

The Noisy-OR Construction

Suppose there are m possible decisions: $|\mathcal{D}| = m$. Our strategy will be to reduce the DRD problem to $O(m)$ instances of the ECD problem, such that solving *any one of them* is sufficient for solving the DRD problem. Crucially, the problem we end up solving depends on the unknown hypothesis h^*. We design our surrogate DIRECT so that it adaptively determines which instance to solve in order to minimize the expected total cost.

Concretely, we construct m different graphs, one for each decision. The role of graph i is to determine whether the unknown hypothesis h^* is contained in decision region $\mathcal{R}_i$ or not. Thus we aim to distinguish all the hypotheses in this decision region from the rest. To achieve this, we model graph i as an ECD problem, with one of the decision regions being $\mathcal{R}_i$. Further, we partition the remaining set of hypotheses $\mathcal{H} \setminus \mathcal{R}_i$ into a collection of *subregions*, such that within each subregion, all hypotheses are contained in exactly the same collection of decision regions from the original DRD problem. All the subregions are disjoint by definition, and hence we have a well-defined ECD problem. Solving this problem amounts to cutting all the edges between $\mathcal{R}_i$ and the subregions. See Figure 1 for illustration.

Notice that in this ECD problem, once all the edges are cut, either i is the optimal decision, or one of the subregions encodes the optimal decision. Therefore, optimizing the ECD problem associated with one of the m graphs is a *sufficient condition* for identifying the optimal decision.

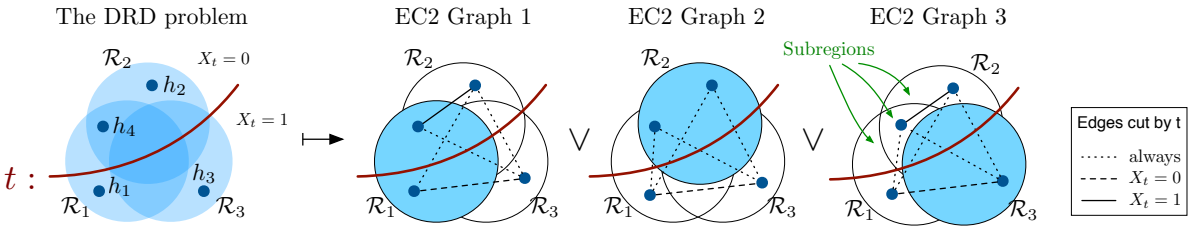


Figure 1: A toy DRD problem with three decision regions $\{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$, and four possible hypotheses $\{h_1, h_2, h_3, h_4\}$. t is a test with two possible outcomes: $f_t(h_1) = f_t(h_3) = 1$ and $f_t(h_2) = f_t(h_4) = 0$. For each possible decision we can make, we construct a separate ECD problem: The three figures on the right illustrate the EC^2 graphs for each of the ECD problems. We can successfully make an optimal decision once one of the graphs is fully cut: e.g., if $X_t = 0$, graph 2 is fully cut, and we identify the optimal decision d_2.

Further notice that, among the m ECD problems associated with the m graphs, at least one of them has to be solved (i.e., all edges cut) before we uncover the optimal decision. Therefore, we get a *necessary condition* of the DRD constraints: we have to cut all the edges in *at least one* of the m graphs. This motives us to apply a logical OR operation on the m optimization problems. Denote the EC^2 objective function for graph i as f^i_{EC}, and normalize them so that $f^i_{EC}(\emptyset) = 0$ corresponds to observing nothing and $f^i_{EC}(\mathbf{x}_\mathcal{T}) = 1$ corresponds to all edges being cut. We combine the objective functions $f^1_{EC}, \ldots, f^m_{EC}$ using a *Noisy-OR formulation*:

$$f_{DRD}(\mathbf{x}_\mathcal{A}) = 1 - \prod_i^m \left(1 - f^i_{EC}(\mathbf{x}_\mathcal{A})\right) \tag{3}$$

Note that by design $f_{DRD}(\mathbf{x}_\mathcal{A}) = 1$ iff $f^i_{EC}(\mathbf{x}_\mathcal{A}) = 1$ for *at least* one i. Thus, the DRD (and hence NVOI-NMU) Problem is formally equivalent to the following problem:

$$\pi^* \in \arg\min_\pi \text{cost}(\pi), \text{ s.t.}$$
$$\forall \mathbf{x}_\mathcal{T} : f_{DRD}(\mathcal{S}(\pi, \mathbf{x}_\mathcal{T})) \geq 1 \text{ whenever } \mathbb{P}[\mathbf{x}_\mathcal{T}] > 0. \tag{4}$$

The crucial advantage of this new formulation is given by the following Lemma:

Lemma 1. *f_{DRD} is strongly adaptive monotone, and adaptive submodular w.r.t. $\mathbb{P}$.*

That is, the Noisy-OR formulation for multiple EC^2 functions preserves adaptive submodularity. The proof of this result can be found in the supplemental material[4]. These properties make f_{DRD} amenable for efficient greedy optimization. Formally, let $\Delta_{f_{DRD}}(t \mid \mathbf{x}_\mathcal{A}) := \mathbb{E}_{x_t}\left[f_{DRD}(\mathbf{x}_{\mathcal{A}\cup\{t\}}) - f_{DRD}(\mathbf{x}_\mathcal{A}) \mid \mathbf{x}_\mathcal{A}\right]$ be the expected marginal benefit in f_{DRD} by adding test t to $\mathbf{x}_\mathcal{A}$. With f_{DRD}, we can associate the following greedy algorithm: It starts with the empty set, and at each iteration, having already observed $\mathbf{x}_\mathcal{A}$, selects the test t^* with the largest benefit-to-cost ratio: $t^* \in \arg\max_t \Delta_{f_{DRD}}(t \mid \mathbf{x}_\mathcal{A})/c(t)$. A major benefit of adaptive submodularity is that we can use a technique called lazy evaluation to dramatically speed up the selection process (Golovin and Krause 2011). Further, we have the following performance guarantee:

Theorem 2. *Let m be the number of decisions, and π_{DRD} be the adaptive greedy policy w.r.t. the objective function Eq.* (3)*. Then it holds that*

$$\text{cost}(\pi_{DRD}) \leq (2m \ln(1/p_{\min}) + 1)\,\text{cost}(\pi^*),$$

where $p_{\min} = \min_{h\in\mathcal{H}} \mathbb{P}[h]$ is the minimum prior probability of any set of observations, and π^* is the optimal policy for Problem (4)*, and hence also the NVOI-NMU and DRD Problems.*

This result follows from Lemma 1 and the general performance analysis of the greedy policy for adaptive submodular problems by (Golovin and Krause 2011). More details are given in the supplement. The bound of the greedy algorithm is linear in the number of decision regions. Here the factor m is a result of taking the product of m EC^2 instances. In the following, we show how this bound can often be improved.

Improving the bound via Graph Coloring

For certain applications, the number of decisions m can be large. In the extreme case where we have a unique decision for each possible observation, the bound of Theorem 2 becomes trivial. As noted, this is a result of taking the product of m EC^2 instances. Thus, we can improve this bound by constructing fewer instances, each with several *non-overlapping* decision regions. As long as every decision region is accounted for by at least one ECD instance, problem 4 remains equivalent to the DRD problem. We select the sets of decision regions for each ECD instance through graph coloring. See Figure 2 for illustration.

Formally, we construct an undirected graph $\mathcal{G} := \{\mathcal{D}, \mathcal{E}\}$ over all decision regions, where we establish an edge between any pair of overlapping decision regions. That is, two decision regions $\mathcal{R}_i$ and $\mathcal{R}_j$ are adjacent in $\mathcal{G}$ iff there exists a hypothesis h for which both decisions are optimal, i.e., $h \in \mathcal{R}_i \cap \mathcal{R}_j$. Finding a minimal set of non-overlapping decision region sets that covers all decisions is equivalent to solving a graph coloring problem, where the goal is to color the vertices of the graph $\mathcal{G}$, such that no two adjacent vertices share the same color, using as few colors as possible. Thus, we can construct one ECD problem for all the decision

[4]Similar constructions have been used for classical submodular set functions (Guillory and Bilmes 2011; Deshpande, Hellerstein, and Kletenik 2014), utilizing the fact that $f = 1 - \prod_i^m (1 - f_i)$ is submodular if each f_i is submodular. However, the function f is *not* necessarily adaptive submodular, even when each f_i is adaptive submodular and strongly adaptively monotone.

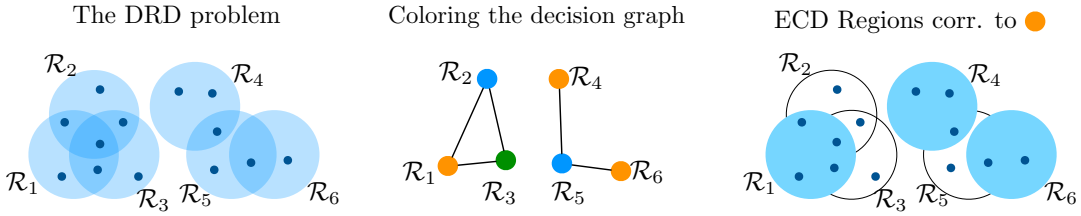


Figure 2: Reducing the cost upper bound via graph coloring. We only need to construct 3 ECD instances to compute f_{DRD}, instead of 6. The middle figure shows a possible coloring assignment on the decision graph of the DRD problem. On the right, we show one example ECD problem instance, corresponding to regions $\{\mathcal{R}_1, \mathcal{R}_4, \mathcal{R}_6\}$ (colored orange). In this ECD problem instance, there are 7 disjoint regions: 3 (disjoint) decision regions $\mathcal{R}_1, \mathcal{R}_4, \mathcal{R}_6$, and 4 subregions, namely $\mathcal{R}_2 \setminus (\mathcal{R}_1 \cup \mathcal{R}_3)$, $\mathcal{R}_3 \setminus (\mathcal{R}_1 \cup \mathcal{R}_2)$, $(\mathcal{R}_2 \cap \mathcal{R}_3) \setminus \mathcal{R}_1$, and $\mathcal{R}_5 \setminus (\mathcal{R}_4 \cup \mathcal{R}_6)$.

regions of the same color, resulting in r different instances, and then use the Noisy-OR formulation to assemble these objective functions. That gives us the following theorem:

Theorem 3. *Let π_{DRD} be the adaptive greedy policy w.r.t. the objective function Eq.* (3)*, which is computed over ECD problem instances obtained via graph coloring. Let r be the number of colors used. Then it holds that*

$$\mathrm{cost}(\pi_{DRD}) \leq (2r \ln(1/p_{\min}) + 1)\,\mathrm{cost}(\pi^*),$$

where $p_{\min}$ is the minimum prior probability of any set of observations, and π^* is the optimal policy.

While obtaining minimum graph colorings is NP-hard in general, one can show that every graph can be efficiently colored with at most one color more than the maximum vertex degree, denoted by $\deg$, using a greedy coloring algorithm (Welsh and Powell 1967): consider the vertices in descending order according to the degree; we assign to a vertex the smallest available color not used by its neighbours, adding a fresh color if needed. In the DRD setting, $\deg$ is the maximal number of decision regions that any decision region can be overlapped with. In practice, greedy coloring often requires far fewer colors than this upper bound. Additionally, note that when regions are disjoint, $\deg = 0$ and DIRECT reverts to the EC^2 algorithm.

Dealing with Noisy Observations

The computational complexity of DIRECT depends *linearly*[5] on the number of hypotheses in the DRD problem, i.e., the number N of all possible outcome vectors $\mathbf{x}_{\mathcal{T}}$ in the NVOI-NMU problem. However, N can be large, in particular in settings where we model complex joint distributions $\mathbb{P}[Y, X_1, \ldots, X_n]$. Fortunately, often one can exploit structure in the probabilistic model to dramatically improve the computation complexity. Note that for any discrete prior $\mathbb{P}[Y, X_1, \ldots, X_n]$ we can define a latent variable Θ, such that $\mathbb{P}[X_1, \ldots, X_n \mid Y, \Theta]$ becomes deterministic, i.e., $X_i = f_i(Y, \Theta)$ for some deterministic function f_i. Thus, we can simply interpret each pair of (y, θ) as a hypothesis h in DRD. One natural reason to introduce the latent variable Θ is to deal with noisy observations. In our medical diagnosis example, patients with the same condition Y may have different symptoms, and thus react differently to the same medical tests. Here Θ captures the possible clinical manifestations that come along with Y.

For some noise models, Θ could have exponentially large support, and thus keeping track of all the noisy realizations of a hypothesis will be prohibitive. To overcome this challenge, we show that for certain distributions of Θ, one can compute the objective f_{DRD} much more efficiently. In particular, we study a natural restricted noise model, where Θ encodes a bounded number of k ($k \ll n$) flips of the ground truth label[6] induced by y. Imagine that for each hidden state y, there is some θ_y such that hypothesis $h = (y, \theta_y)$ corresponds to the "clean" state of y, while hypotheses $\hat{h} \in \{(y, \hat{\theta}_y) : \hat{\theta}_y \neq \theta_y\}$ correspond to the noisy versions of y. We further suppose that a noisy hypothesis $\hat{h}$ flips the label of each test with probability ϵ, and that the total number of label flips, denoted as $\delta(h, \hat{h})$, follows a truncated binomial distribution: $\mathbb{P}[\delta] = \frac{1}{Z}\binom{n}{\delta}\epsilon^{\delta}(1-\epsilon)^{n-\delta}$ for $\delta \leq k$ (and $\mathbb{P}[\delta] = 0$ for $\delta > k$), where $Z = \sum_{\delta=1}^{k}\binom{n}{\delta}\epsilon^{d}(1-\epsilon)^{n-\delta}$ is the normalizing constant. Since the utility function in the NVOI-NMU problem is defined in terms of the state y, to compute f_{DRD}, it suffices to be able to efficiently compute the remaining probability mass $\mathbb{P}[y, \mathbf{x}_{\mathcal{A}}]$ associated with state y after observing $\mathbf{x}_{\mathcal{A}}$. Let $l = |\{t : t \in \mathcal{A} \wedge f_t(h) \neq \mathbf{x}_t\}|$ be the number of labels in $\mathbf{x}_{\mathcal{A}}$ that are inconsistent with (y, θ_y). One can show that $\mathbb{P}[y, \mathbf{x}_{\mathcal{A}}] = \mathbb{P}[y] \cdot \frac{1}{Z}\epsilon^{l}(1-\epsilon)^{|\mathcal{A}|-l}\sum_{i=0}^{k-l}\binom{n-|\mathcal{A}|}{i}\epsilon^{i}(1-\epsilon)^{n-|\mathcal{A}|-i}$. Therefore, $\mathbb{P}[y, \mathbf{x}_{\mathcal{A}}]$ can be computed efficiently *without* enumerating all $O\left(n^k\right)$ (y, θ) pairs, and hence we can compute f_{DRD} efficiently as well.

Experimental Results

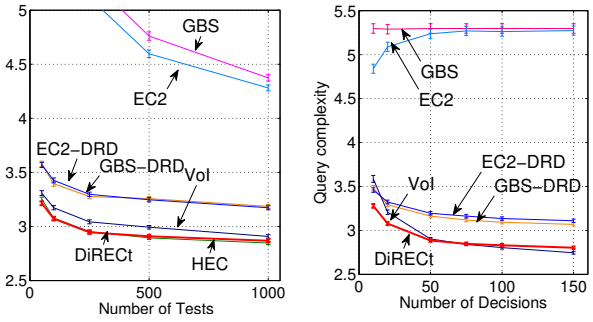
We now consider four instances of the general non-myopic value of information problem. Table 1 summarizes how these instances fit into our framework. For each of the problems, we compare DIRECT against several existing approaches as baselines. The first baseline is myopic optimization of the decision-theoretic value of information (VOI) (Howard 1966). At each step we greedily choose the test that maximizes the expected value given the current observations $\mathbf{x}_{\mathcal{A}}$, i.e., $t \in \arg\max_t \mathbb{E}_{x_t}\left[U(\mathbf{x}_{\mathcal{A}\cup\{x\}})\right]$. The second baseline is the recently proposed objective for addressing the DRD problem, HEC (Javdani et al. 2014). We also compare

[5] Since DIRECT requires the computation of r EC^2 scores, the computational complexity of DIRECT is *linear* in both r and N.

[6] We assume, w.l.o.g., that test outcomes are binary.

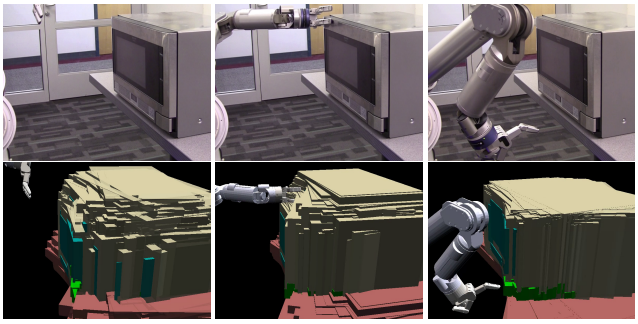
APPLICATION	TEST / ACTION	DECISION
Active Loc.	guarded move	manipulation action
Pref. learning.	pair of movies	recommendation
Conservation	monitoring / probing	conservation action
Risky choice	pair of lottery choices	valuation theory

Table 1: Tests and decisions for different applications



(a) *Robot* - varying $|\mathcal{T}|$ (b) *Robot* - varying $|\mathcal{D}|$

Figure 3: Experimental results - *Robot*



(a) Hypotheses (b) Tests (c) Decision regions

Figure 4: Experimental setup for touch-based localization. (a) Uncertainty is represented by hypotheses over object pose. (b) Tests are guarded moves, where the end effector moves along a path until contact is sensed. Hypotheses which could not have produced contact at that location (e.g. they are too far or too close) are removed. (c) Decisions are button-push attempts: trajectories starting at a particular location, and moving forward. The corresponding region consists of all poses for which that button push would succeed.

with algorithms designed for special cases of the DRD problem: generalized binary search (GBS) and equivalence class edge cutting (EC^2)[7]. We compare with two versions of these algorithms: one with the algorithms' original stopping criteria, which we call GBS and EC^2; and one with the stopping criteria of the DRD problem, which is referred to as GBS-DRD and EC^2-DRD in the results.

Active touch-based localization

Our first application is a robotic manipulation task of pushing a button, with uncertainty over the target's pose. Tests consist of *guarded moves* (Will and Grossman 1975), where the end effector moves along a path until contact is sensed. Those hypotheses which would not have produced contact at that location (e.g., they are far away) can be eliminated. Decisions correspond to putting the end effector at a particular location and moving forward. The coinciding decision region consists of all object poses where the button would successfully be pushed. Our goal is to concentrate all consistent hypotheses within a single decision region using the fewest tests. We model pose uncertainty with 4 parameters: (x, y, z) for positional uncertainty, and θ for rotation about the z axis. An initial set of 20000 hypotheses are sampled from a normal distribution $N(\mu, \Sigma)$, where μ is some initial location (e.g., from a camera), and Σ is diagonal with $\sigma_x = \sigma_y = \sigma_z = 2.5$cm, and $\sigma_\theta = 7.5°$. We then run DIRECT on both simulated data and a real robot platform. In the first simulated experiment, we preselect a grid of 25 button pushing actions $\mathcal{D}$ while ensuring the overlap r is minimal. We randomly generate guarded moves $\mathcal{T}$ to select from, varying $|\mathcal{T}|$. In the second, we randomly generate decision regions, varying $|\mathcal{D}|$ while fixing $|\mathcal{T}| = 250$. To compute the *myopic value of information* (VOI) (Howard 1966), we define a utility function $u(h, \mathcal{R})$ which is 1 if $h \in \mathcal{R}$ and 0 otherwise. Results are plotted in Figure 3(a) and Figure 3(b). Note that HEC cannot be computed in this experiment, as the overlap r becomes very large[8] and HEC quickly becomes intractable. We see that DIRECT generally outperforms other baselines. Here, myopic VOI performs comparably – likely because the problem is solved within a short horizon. We also demonstrate DIRECT on a real robot platform as illustrated in Figure 4. See supplemental material for more results and a video demonstration.

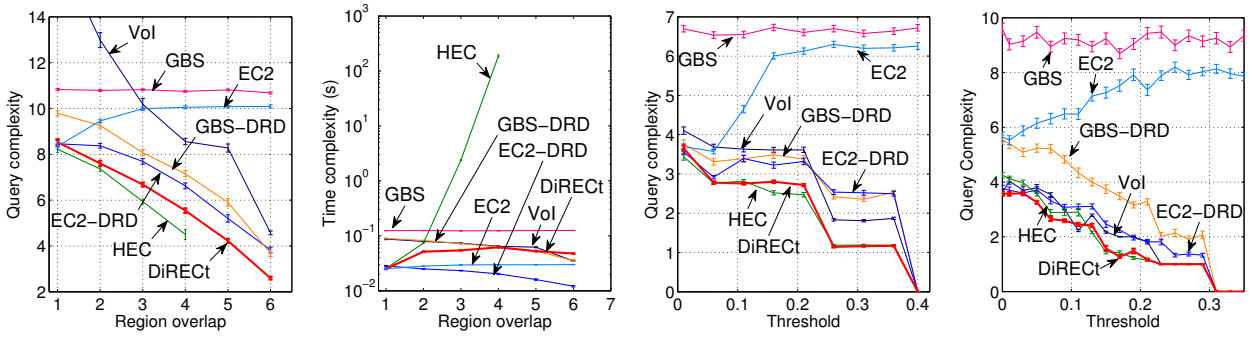
Comparison-based preference learning

The second application considers a comparison-based movie recommendation system, which learns a user's movie preference (e.g., the favorable genre) by sequentially showing her pairs of candidate movies, and letting her choose which one she prefers. We use the *MovieLens 100k* dataset (Herlocker et al. 1999), which consists a matrix of 1 to 5 ratings of 1682 movies from 943 users. For fair comparison with baselines, we adopt the same parameters as reported in (Javdani et al. 2014). That is, for each movie we extract a 10-d feature representation from the rating matrix through SVD. To generate decision regions, we cluster movies using k-means, and assign each movie to the r closest cluster centers.

We demonstrate the performance of DIRECT on *MovieLens* in Figure 5(a) and 5(b). We fix the number of clusters (i.e., decision regions) to 12, and vary r, the number of assigned regions for each hypothesis, from 1 to 6. Note that r controls the hyperedge cardinality in HEC, which crucially

[7]When hypotheses are in multiple decision regions, EC^2 cannot be used as is. Hence, we randomly assign each hypothesis to one of the decision regions that it is contained in.

[8]Moreover, when running on a real robot, many actions are infeasible due to kinematic constraints. Sampling decisions enables us to generate arbitrarily many, ensuring we always have many decisions available.



(a) *MovieLens* (b) *MovieLens* - time (c) *EMPCranes* - Allow 1 flip (d) *Risky Choice Theory*

Figure 5: Experimental results: *MovieLens*, *EMPCranes*, and *Risky Choice Theory*

affects the computational complexity. As we can observe, the *query complexity* (i.e., the number of queries needed to identify the target region) of DIRECT is lower than all baselines except HEC. However, it is significantly faster to compute. See Figure 5(b) (for $r = 5$, HEC failed to pick any tests within an hour).

Adaptive management for wild-life conservation

Our third application is a real-world value of information problem in natural resource management, where one needs to determine which management action should be undertaken for wild-life conservation. Specifically, the task is to preserve the *Eastern Migration Population of whooping cranes* (*EMP Cranes*). An expert panel came up with 8 hypotheses for possible causes of reproductive failure, along with 7 management strategies (as decisions). The decision-hypothesis utility matrix is specified in Table 5 of (Runge, Converse, and Lyons 2011). Tests aim to resolve specific sources of uncertainty. Our goal is to find the best conservation strategy using the minimal number of tests.

We assume that ε-optimal decisions are allowed for each hypothesis, where ε is the tolerance threshold. We further assume test outcomes to be noisy, i.e., the test outcome corresponding to a particular hypothesis can be flipped. In our experiments, a maximum of 1 flip is allowed for each outcome vector, which amounts to a total of 37 "noisy" hypotheses. When multiple hypotheses are consistent with a outcome vector, we assign the most probable one to that outcome. Results are plotted in Figure 5(c). We see that HEC and DIRECT perform comparably well, while significantly outperforming myopic VOI and all other baselines.

Preference Elicitation in Behavioral Economics

We further conduct experiments in an experimental design task. Several theories have been proposed in behavioral economics to explain how people make decisions under risk and uncertainty. We test DIRECT on six theories of subjective valuation of risky choices (Wakker 2010; Tversky and Kahneman 1992; Sharpe 1964), namely the (1) *expected utility with constant relative risk aversion*, (2) *expected value*, (3) *prospect theory*, (4) *cumulative prospect theory*, (5) *weighted moments*, and (6) *weighted standardized moments*. Choices are between risky lotteries, i.e., known distribution over payoffs (e.g., the monetary value gained or lost). Tests are pairs of lotteries, and hypotheses correspond to parametrized theories that predict, for a given test, which lottery is preferable. The goal, is to adaptively select a sequence of tests to present to a human subject in order to distinguish which of the six theories best explains the subject's responses.

We employ the same set of parameters used in (Ray et al. 2012) to generate tests and hypotheses. The original setup in (Ray et al. 2012) was designed for testing EC^2, and therefore test realizations of different theories cannot collide. In our experiments, we allow a tolerance ϵ - that is, if one hypothesis differs from another by at most ε, they are considered to be similar, and thus have the same set of optimal decisions. Results for simulated test outcomes with varying ε are shown in Figure 5(d). We see that DIRECT performs best in this setting.

Conclusion

We have proposed DIRECT, an efficient surrogate for the problem of nonmyopically optimizing value of information to achieve near-maximal utility. We prove that DIRECT is adaptive submodular, making it amenable for efficient greedy optimization. We demonstrated the efficiency and effectiveness of DIRECT extensively on four real-world applications, and showed that it compares favorably with existing approaches, while being significantly faster than competing methods. We believe that our results provide an important step towards solving challenging real-world information gathering problems.

Acknowledgements. This work was supported in part by the Intel Embedded Computing ISTC, NSF NRI Purposeful Prediction grant, NSF GRFP No. 0946825, NSF-IIS-1227495, DARPA MSEE FA8650-11-1-7156, ERC StG 307036, a Microsoft Research Faculty Fellowship, a Google European Doctoral Fellowship, and the Office of Naval Research Young Investigator Award.

References

Balcan, M.; Beygelzimer, A.; and Langford, J. 2006. Agnostic active learning. In *ICML*.

Chakaravarthy, V. T.; Pandit, V.; Roy, S.; Awasthi, P.; and Mohania, M. 2007. Decision trees for entity identification: Approximation algorithms and hardness results. In *SIGMOD/PODS*.

Chaloner, K., and Verdinelli, I. 1995. Bayesian experimental design: A review. *Statistical Science* 10(3):273–304.

Chen, S. J.; Choi, A.; and Darwiche, A. 2014. Algorithms and applications for the same-decision probability. *JAIR* 601–633.

Choi, A.; Xue, Y.; and Darwiche, A. 2012. Same-decision probability: A confidence measure for threshold-based decisions. *International Journal of Approximate Reasoning* 53(9):1415 – 1428.

Dasgupta, S. 2004. Analysis of a greedy active learning strategy. In *NIPS*.

Deshpande, A.; Hellerstein, L.; and Kletenik, D. 2014. Approximation algorithms for stochastic boolean function evaluation and stochastic submodular set cover. In *SODA*.

Golovin, D., and Krause, A. 2011. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *JAIR*.

Golovin, D.; Krause, A.; and Ray, D. 2010. Near-optimal bayesian active learning with noisy observations. *CoRR*.

Guillory, A., and Bilmes, J. 2011. Simultaneous learning and covering with adversarial noise. In *ICML*, 369–376.

Hanneke, S. 2007. A bound on the label complexity of agnostic active learning. In *ICML*.

Herlocker, J. L.; Konstan, J. A.; Borchers, A.; and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *SIGIR*.

Howard, R. A. 1966. Information value theory. In *IEEE Transactions on Systems Science and Cybernetics*.

Javdani, S.; Klingensmith, M.; Bagnell, J. A. D.; Pollard, N.; and Srinivasa, S. 2013. Efficient touch based localization through submodularity. In *ICRA*.

Javdani, S.; Chen, Y.; Karbasi, A.; Krause, A.; Bagnell, D.; and Srinivasa, S. 2014. Near-optimal bayesian active learning for decision making. In *AISTATS*.

Kaelbling, L. P.; Littman, M. L.; and Cassandra, A. R. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101:99–134.

Karbasi, A.; Ioannidis, S.; and Massoulié, L. 2011. Content search through comparisons. In *Automata, Languages and Programming*, volume 6756 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 601–612.

Karbasi, A.; Ioannidis, S.; and Massoulié, L. 2012. Comparison-based learning with rank nets. In *ICML*.

Kononenko, I. 2001. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine* 23:89–109.

Kosaraju, S. R.; Przytycka, T. M.; and Borgstrom, R. S. 1999. On an optimal split tree problem. In *WADS'99*, 157–168.

Krause, A., and Guestrin, C. 2009. Optimal value of information in graphical models. *JAIR* 35:557–591.

Pineau, J.; Gordon, G.; and Thrun, S. 2006. Anytime point-based approximations for large pomdps. *JAIR* 27(1):335–380.

Ray, D.; Golovin, D.; Krause, A.; and Camerer, C. 2012. Bayesian rapid optimal adaptive design (broad): Method and application distinguishing models of risky choice. *Tech. Report*.

Runge, M. C.; Converse, S. J.; and Lyons, J. E. 2011. Which uncertainty? using expert elicitation and expected value of information to design an adaptive program. *Biological Conservation*.

Settles, B. 2012. *Active Learning*. Morgan & Claypool.

Sharpe, W. F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*.

Smallwood, R. D., and Sondik, E. J. 1973. The optimal control of partially observable markov processes over a finite horizon. *Operations Research* 21(5):1071–1088.

Tversky, A., and Kahneman, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4).

Wakker, P. 2010. *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.

Welsh, D. J., and Powell, M. B. 1967. An upper bound for the chromatic number of a graph and its application to timetabling problems. *Computer Journal*.

Will, P. M., and Grossman, D. D. 1975. An experimental system for computer controlled mechanical assembly. *IEEE Trans. Computers* 24(9):879–888.