
Interactive Submodular Set Cover

Andrew Guillory

Jeff Bilmes

University of Washington, Seattle, WA 98195, USA

GUILLORY@CS.WASHINGTON.EDU

BILMES@EE.WASHINGTON.EDU

Abstract

We introduce a natural generalization of submodular set cover and exact active learning with a finite hypothesis class (query learning). We call this new problem interactive submodular set cover. Applications include advertising in social networks with hidden information. We give an approximation guarantee for a novel greedy algorithm and give a hardness of approximation result which matches up to constant factors. We also discuss negative results for simpler approaches and present encouraging early experimental results.

1. Introduction

As a motivating example, we consider viral marketing in a social network. In the standard version of the problem, the goal is to send advertisements to influential members of a social network such that by sending advertisements to only a few people our message spreads to a large portion of the network. Previous work (Kempe et al., 2005; 2003) has shown that, for many models of influence, the influence of a set of nodes can be modelled as a submodular set function. Therefore, selecting a small set of nodes with maximal influence can be posed as a *submodular function maximization* problem. The related problem of selecting a minimal set of nodes to achieve a desired influence is a *submodular set cover* problem. Both of these problems can be approximately solved via a simple greedy approximation algorithm.

Consider a variation of this problem in which the goal is not to send advertisements to people that are influential in the entire social network but rather to people that are influential in a specific target group. For example, our target group could be people that like

snowboarding or people that listen to jazz music. If the members of the target group are unknown and we have no way of learning the members of the target group, there is little we can do except assume every member of the social network is a member of the target group. However, if we assume the group has some known structure and that we receive feedback from sending advertisements (e.g. in the form of ad clicks or survey responses), it may be possible to simultaneously discover the members of the group and find people that are influential in the group.

We call problems like this *learning and covering problems*. In our example, the *learning* aspect of the problem is discovering the members of the target group (the people that like snowboarding), and the *covering* aspect of the problem is to select a small set of people that achieve a desired level of influence in the target group (the people to target with advertisements). Other applications have similar structure. For example, we may want to select a small set of representative documents about a topic of interest (e.g. about linear algebra). If we do not initially know topic labels for documents, this is a learning and covering problem.

We propose a new problem called *interactive submodular set cover* that can be used to model many learning and covering problems. Besides addressing interesting new applications, interactive submodular set cover directly generalizes submodular set cover and exact active learning with a finite hypothesis class (query learning) giving new insight into many previous theoretical results. We derive and analyze a new algorithm that is guaranteed to perform approximately as well as any other algorithm. Our algorithm considers simultaneously the learning and covering parts of the problem. It is tempting to try to treat these two parts of the problem separately for example by first solving the learning problem and then solving the covering problem. We prove this approach and other simple approaches may perform much worse than the optimal algorithm. Some proofs are omitted for space but are in the technical report (Guillory & Bilmes, 2010).

Appearing in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, 2010. Copyright 2010 by the author(s)/owner(s).

2. Background

2.1. Submodular Set Cover

A submodular function is a set function satisfying a natural diminishing returns property. We call a set function F defined over a ground set V submodular iff for all $A \subseteq B \subseteq V$ and $v \in V \setminus B$

$$F(A + v) - F(A) \geq F(B + v) - F(B) \quad (1)$$

In other words, adding an element to A , a subset of B , results in a larger gain than adding the same element to B . F is called modular if Equation 1 holds with equality. F is monotone non-decreasing if for all $A \subseteq B \subseteq V$, $F(A) \leq F(B)$.

In the submodular set cover problem the goal is to find a set $S \subseteq V$ minimizing a modular cost function $c(S) = \sum_{s \in S} c(s)$ subject to the constraint $F(S) = F(V)$ for a monotone non-decreasing submodular F .

This problem is closely related to the problem of submodular function maximization under a modular cost constraint $c(S) < k$ for a constant k . A number of interesting real world applications can be posed as submodular set cover or submodular function maximization problems including influence maximization in social networks (Kempe et al., 2003), sensor placement and experiment design (Krause et al., 2008), and document summarization (Lin & Bilmes, 2010). In the sensor placement problem, for example, the ground set V corresponds to a set of possible locations. An objective function $F(S)$ measures the coverage achieved by deploying sensors to the locations corresponding to $S \subseteq V$. For many reasonable definitions of coverage, $F(S)$ turns out to be submodular.

Submodular set cover is a generalization of the set cover problem, and, as is the case for set cover, a greedy algorithm has approximation guarantees (Wolsey, 1982). In particular, if F is integer valued, then the greedy solution is within $H(\max_{v \in V} F(\{v\}))$ of the optimal solution where $H(k)$ is the k th harmonic number. Up to lower order terms, this matches the hardness of approximation lower bound for set cover (Feige, 1998). We note a variation of submodular set cover uses a constraint $F(S) \geq \alpha$ for a fixed threshold α . It can be shown this variation is equivalent (Krause et al., 2008; Narayanan, 1997), so without loss of generality or specificity, we use this variation.

2.2. Exact Active Learning

In the exact active learning problem we have a known finite hypothesis class given by a set of objects H , and we want to identify an initially unknown target

hypothesis $h^* \in H$. We identify h^* by asking questions. Define Q to be the known set of all possible questions. A question q maps an object h to a set of valid responses $q(h) \subseteq R$ with $q(h) \neq \emptyset$ where $R \triangleq \bigcup_{q \in Q, h \in H} q(h)$ is the set of all possible responses. We know the mapping for each q (i.e. we know $q(h)$ for every q and h). Asking q reveals some element $r \in q(h^*)$ which may be chosen adversarially (chosen to impede the learning algorithm). Each question $q \in Q$ has a positive cost $c(q)$ defined by the modular cost function c .

The goal of active learning is to ask a sequence of questions with small total cost that identifies h^* . By identifying h^* , we mean that for every $h \neq h^*$ we have received some response r to a question q such that $r \notin q(h)$. Questions are chosen sequentially so that the response from a previous question can be used to decide which question to ask next.

In a typical exact learning problem, H is a set of different classifiers and h^* is a unique zero-error classifier. Questions in Q can, for example, correspond to label (membership) queries for data points. If we have a fixed data set consisting of data points x_i , we can create a question q_i corresponding to each x_i and set $q_i(h) = \{h(x_i)\}$. We note that we assume that there is no noise in responses and that the hypothesis class is correct ($h^* \in H$). Questions can also correspond to more complicated queries. For example, a question can ask if any points in a set are positively labelled. The setting we have described allows for mixing arbitrary types of queries with different costs.

For a set of question-response pairs \hat{S} , define the version space $V(\hat{S})$ to be the subset of H consistent with \hat{S} . More formally $V(\hat{S}) \triangleq \{h \in H : \forall (q, r) \in \hat{S}, r \in q(h)\}$. In terms of the version space, the goal of exact active learning is to ask a sequence of questions such that $|V(\hat{S})| = 1$.

Building on previous work (Balcázar et al., 2007), Hanneke (2006) showed that a simple greedy active learning strategy is approximately optimal in the setting we have described. The greedy strategy selects the question which relative to cost distinguishes the greatest number of hypotheses from h^* . Hanneke (2006) shows this strategy incurs no more than $\ln |H|$ times the cost of any other question asking strategy.

3. Problem Statement

We use notation similar to the exact active learning problem we described in the previous section. Assume we have a finite hypothesis class H containing an unknown target hypothesis $h^* \in H$. We again assume

there is a finite set of questions Q , a question q maps each object h to a set of valid responses $q(h) \subseteq R$ with $q(h) \neq \emptyset$, and each question $q \in Q$ has a positive cost $c(q)$ defined by the modular cost function c . We also again assume that we know the mapping for each q (i.e. we know $q(h)$ for every q and h). Asking q reveals some adversarially chosen element $r \in q(h^*)$. In the exact active learning problem the goal is to identify h^* through questions. In this work we consider a generalization in which the goal is instead to satisfy a submodular constraint that depends on h^* .

We assume that for each object h there is a corresponding monotone non-decreasing submodular function F_h defined over subsets of $Q \times R$ (sets of question-response pairs). We repeatedly ask a question \hat{q}_i and receive a response \hat{r}_i . Let the sequence of questions be $\hat{Q} = (\hat{q}_1, \hat{q}_2, \dots)$ and sequence of responses be $\hat{R} = (\hat{r}_1, \hat{r}_2, \dots)$. Define $\hat{S} = \bigcup_{\hat{q}_i \in \hat{Q}} \{(\hat{q}_i, \hat{r}_i)\}$ to be the final set of question-response pairs corresponding to these sequences. Our goal is to ask a sequence of questions with minimal total cost $c(\hat{Q})$ so $F_{h^*}(\hat{S}) \geq \alpha$ for some threshold α without knowing h^* beforehand. We call this interactive submodular set cover.

Interactive Submodular Set Cover

Given:

- Hypothesis class H containing an unknown target h^*
- Query set Q and response set R with known $q(h) \subseteq R$ for every $q \in Q, h \in H$
- Modular query cost function c defined over Q
- Submodular monotone non-decreasing objective functions F_h for $h \in H$ defined over $Q \times R$
- Objective threshold α

Repeat: Ask a question $\hat{q}_i \in Q$ and receive a response $\hat{r}_i \in \hat{q}_i(h^*)$

Until: $F_{h^*}(\hat{S}) \geq \alpha$ where $\hat{S} = \bigcup_i \{(\hat{q}_i, \hat{r}_i)\}$

Objective: Minimize $\sum_i c(\hat{q}_i)$

Note that although we know the hypothesis class H and the corresponding objective functions F_h , we do not initially know h^* . Information about h^* is only revealed as we ask questions and receive responses to questions. Responses to previous questions can be used to decide which question to ask next, so in this way the problem is “interactive.” Furthermore, the objective function for each hypothesis F_h is defined over sets of question-response pairs (as opposed to, say, sets of questions), so when asking a new question we cannot predict how the value of F_h will change until after we receive a response. The only restriction on the response we receive is that it must be consistent with the initially unknown target h^* . It is this uncertainty

about h^* and the feedback we receive from questions that allows us to model learning and covering.

3.1. Connection to Submodular Set Cover

If we know h^* (e.g. if $|H| = 1$) and we assume $|q(h)| = 1 \forall q \in Q, h \in H$ (i.e. that there is only one valid response to every question), our problem reduces exactly to the standard submodular set cover problem. Under these assumptions, we can compute $F_{h^*}(\hat{S})$ for any set of questions without actually asking these questions. Krause et al. (2008) study a non-interactive version of interactive submodular set cover in which $|q(h)| = 1 \forall q \in Q, h \in H$ and the entire sequence of questions must be chosen before receiving any responses. This restricted version of the problem can also be reduced to standard submodular set cover Krause et al. (2008).

3.2. Connection to Active Learning

Define

$$F_h(\hat{S}) \triangleq F(\hat{S}) = |H \setminus V(\hat{S})|$$

where $V(\hat{S})$ is again the version space (the set of hypotheses consistent with \hat{S}). This objective is the number of hypotheses eliminated from the version space by \hat{S} and is submodular and monotone non-decreasing.

For this objective, if we set $\alpha = |H| - 1$ we get the standard exact active learning problem: our goal is to identify h^* using a set of questions with small total cost. Note that in this case the objective F_h does not actually depend on h (i.e. $F_h = F_{h'}$ for all $h, h' \in H$) but the problem still differs from standard submodular set cover because $F_h(\hat{S})$ is defined over question-response pairs.

We can also model other variations of active learning; for example, instead of requiring $|V(\hat{S})| = 1$, with an appropriate F_h we can require that every $h \in V(\hat{S})$ has low error with respect to h^* .

3.3. Connection to Adaptive Submodularity

In concurrent work, Golovin & Krause (2010) show results similar to ours for a different but related class of problems which also involve interactive (i.e. sequential, adaptive) optimization of submodular functions. What Golovin & Krause call realizations correspond to hypotheses in our work while items and states correspond to queries and responses respectively. Golovin & Krause consider both average-case and worst-case settings and both maximization and min-cost coverage problems. In contrast, we only consider worst-case, min-cost coverage problems. In this sense our results are less general.

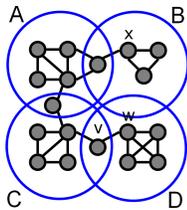


Figure 1. A cartoon example social network.

However, in other ways our results are more general. The main greedy approximation guarantees shown by Golovin & Krause require that the problem is *adaptive submodular*; adaptive submodularity depends not only on the objective but also on the set of possible realizations and the probability distribution over these realizations. In contrast we only require that for a fixed hypothesis the objective is submodular. Golovin & Krause call this pointwise submodularity. Pointwise submodularity does not in general imply adaptive submodularity (see the clustered failure model discussed by Golovin & Krause).

Some other previous work also considers interactive versions of covering problems in the average-case (Asadpour et al., 2008; Goemans & Vondrák, 2006).

4. Example

In the advertising application we described in the introduction, the target hypothesis h^* corresponds to the group of people we want to target with advertisements (e.g. the people that like snowboarding), and the hypothesis class H encodes our prior knowledge about h^* . For example, if we know the target group forms a small dense subgraph in the social network, then the hypothesis class H would be the set of *all* small dense subgraphs in the social network. The query set Q and response set R correspond to advertising actions and feedback respectively, and finally the objective function F_h measures advertising coverage within the group corresponding to h .

To make the discussion concrete, assume the advertiser sends a single ad at a time and that after a person is sent an ad the advertiser receives a binary response indicating if that person is in the target group (i.e. likes snowboarding). Let q_i correspond to sending an ad to user i (i.e. node i), and $q_i(h) = \{1\}$ if user i is in group h and $q_i(h) = \{0\}$ otherwise. For our coverage goal, assume the advertiser wants to ensure that every person in the target group either receives an ad or has a friend that receives an ad. We say a node is “covered” if it has received an ad or has a neighbor that has received

an ad. This is a variation of the minimum dominating set problem, and we use the following objective

$$F_h(\hat{S}) \triangleq \sum_{v \in V_h} I(v \in V_{\hat{S}} \text{ or } \exists s \in V_{\hat{S}} : (v, s) \in E) + |V \setminus V_h|$$

where V and E are the nodes and edges in the social network, V_h is the set of nodes in group h , and $V_{\hat{S}}$ is the set of nodes corresponding to ads we have sent. With this objective $F_{h^*}(\hat{S}) = |V|$ iff we have achieved the stated coverage goal. This objective is also submodular and monotone non-decreasing.

Figure 1 shows a cartoon social network. For this example, assume the advertiser knows the target group is one of the four clusters shown (marked A, B, C, and D) but does not know which. This is our hypothesis class H . The node marked v is initially very useful for learning the members of the target group: if we send an ad to this node, no matter what response we receive we are guaranteed to eliminate two of the four clusters (either A and B or C and D). However, this node has only a degree of 2 and therefore sending an ad to this node does not cover very many nodes. On the other hand, the nodes marked x and w are connected to every node in clusters B and D respectively. x (resp. w) is therefore very useful for achieving the coverage objective if the target group is B (resp. D). An algorithm for learning and covering must choose between actions more beneficial for learning vs. actions more beneficial for covering (although sometimes an action can be beneficial for both to a certain degree). The interplay between learning and covering is similar to the exploration-exploitation trade-off in reinforcement learning. In this example an optimal strategy is to end an ad to v and then cover the remaining two clusters using two ads for a worst case cost of 3.

5. Greedy Approximation Guarantee

We are interested in approximately optimal polynomial time algorithms for the interactive submodular set cover problem. We call a question asking strategy correct if it always asks a sequence of questions such that $F_{h^*}(\hat{S}) \geq \alpha$ where \hat{S} is again the final set of question-response pairs. A necessary and sufficient condition to ensure $F_{h^*}(\hat{S}) \geq \alpha$ for worst case choice of h^* is to ensure $\min_{h \in V(\hat{S})} F_h(\hat{S}) \geq \alpha$ where $V(\hat{S})$ is the version space. Then a simple stopping condition which ensures a question asking strategy is correct is to continue asking questions until $\min_{h \in V(\hat{S})} F_h(\hat{S}) \geq \alpha$. We call a question asking strategy approximately optimal if it is correct and the worst case cost incurred by the strategy is not much worse than the worst case

Algorithm 1 Worst Case Greedy

```

1:  $\hat{H} \leftarrow H$ 
2:  $\hat{S} \leftarrow \emptyset$ 
3: while  $\bar{F}_\alpha(\hat{S}) < \alpha$  do
4:    $\hat{q} \leftarrow \operatorname{argmax}_{q_i \in Q} \min_{h \in V(\hat{S})} \min_{r_i \in q_i(h)} (\bar{F}_\alpha(\hat{S} + (q_i, r_i)) - \bar{F}_\alpha(\hat{S}))/c(q_i)$ 
5:   Ask  $\hat{q}$  and receive response  $\hat{r}$ 
6:    $\hat{S} \leftarrow \hat{S} + (\hat{q}, \hat{r})$ 
7: end while
    
```

cost of any other strategy.

As discussed informally in the previous section, it is important for a question asking strategy to balance between learning (identifying h^*) and covering (increasing F_{h^*}). Ignoring either aspect of the problem is in general suboptimal (we show this formally in Section 6). We propose a reduction which converts the problem over many objective functions F_h into a problem over a single objective function \bar{F}_α that encodes the trade-off between learning and covering. We can then use a greedy algorithm to maximize this single objective, and this turns out to overcome the shortcomings of simpler approaches. This reduction is inspired by the reduction used by Krause et al. (2008) in the non-interactive setting to convert multiple covering constraints into a single covering constraint.

Define

$$\bar{F}_\alpha(\hat{S}) \triangleq (1/|H|) \left(\sum_{h \in V(\hat{S})} \min(\alpha, F_h(\hat{S})) + \alpha |H \setminus V(\hat{S})| \right)$$

$\bar{F}_\alpha(\hat{S}) \geq \alpha$ iff $F_h(\hat{S}) \geq \alpha$ for all $h \in V(\hat{S})$ so a question asking strategy is correct iff it satisfies $\bar{F}_\alpha(\hat{S}) \geq \alpha$. This objective balances the value of learning and covering. The sum over $h \in V(\hat{S})$ measures progress towards satisfying the covering constraint for hypotheses h in the current version space (covering). The second term $\alpha |H \setminus V(\hat{S})|$ measures progress towards identifying h^* through reduction in version space size (learning). Note that the objective does not make a hard distinction between learning actions and covering actions. In fact, the objective will prefer actions that both increase $F_h(\hat{S})$ for $h \in V(\hat{S})$ and decrease the size of $V(\hat{S})$. Crucially, \bar{F}_α retains submodularity.

Lemma 1. *\bar{F}_α is submodular and monotone non-decreasing when every F_h is submodular and monotone non-decreasing.*

Algorithm 1 shows the worst case greedy algorithm which at each step picks the question q_i that maximizes the worst case gain of \bar{F}_α

$$\min_{h \in V(\hat{S})} \min_{r_i \in q_i(h)} (\bar{F}_\alpha(\hat{S} + (q_i, r_i)) - \bar{F}_\alpha(\hat{S}))/c(q_i)$$

We now argue that Algorithm 1 is an approximately optimal algorithm for interactive submodular set cover. Note that although it is a simple greedy algorithm over a single submodular objective, the standard submodular set cover analysis doesn't apply: the objective function is defined over question-response pairs, and the algorithm cannot predict the actual objective function gain until after selecting and committing to a question and receiving a response. We use an Extended Teaching Dimension style analysis (Hanneke, 2006) inspired by previous work in query learning. We are the first to our knowledge to use this kind of proof for a submodular optimization problem.

Define an oracle (teacher) $T \in R^Q$ to be a function mapping questions to responses. As a short hand, for a sequence of questions \hat{Q} define

$$T(\hat{Q}) \triangleq \bigcup_{\hat{q}_i \in \hat{Q}} \{(\hat{q}_i, T(\hat{q}_i))\}$$

$T(\hat{Q})$ is the set of question-response pairs received when T is used to answer the questions in \hat{Q} . We now define a quantity analogous to the General Identification Cost for exact active learning (Hanneke, 2006). Define the General Cover Cost, GCC

$$GCC \triangleq \max_{T \in R^Q} \left(\min_{\hat{Q}: \bar{F}_\alpha(T(\hat{Q})) \geq \alpha} c(\hat{Q}) \right)$$

GCC depends on H , Q , α , c , and the objective functions F_h , but for simplicity of notation this dependence is suppressed. GCC can be viewed as the cost of satisfying $\bar{F}_\alpha(T(\hat{Q})) \geq \alpha$ for worst case choice of T where the choice of T is *known to the algorithm selecting \hat{Q}* . Here the worst case choice of T is over all mappings between Q and R . There is no restriction that T answer questions in a manner consistent with any hypothesis $h \in H$.

We first show that GCC is a lower bound on the optimal worst case cost of satisfying $F_{h^*}(\hat{S}) \geq \alpha$.

Lemma 2. *If there is a correct question asking strategy for satisfying $F_{h^*}(\hat{S}) \geq \alpha$ with worst case cost C^* then $GCC \leq C^*$.*

Proof. Assume the lemma is false and there is a correct question asking strategy with worst case cost C^* and $GCC > C^*$. Using this assumption and the definition of GCC , there is some oracle T^* such that

$$\min_{\hat{Q}: \bar{F}_\alpha(T^*(\hat{Q})) \geq \alpha} c(\hat{Q}) = GCC > C^*$$

When we use T^* to answer questions, any sequence of questions \hat{Q} with total cost less than or equal to

C^* must have $\bar{F}_\alpha(\hat{S}) < \alpha$. $\bar{F}_\alpha(\hat{S}) < \alpha$ in turn implies $F_{h^*}(\hat{S}) < \alpha$ for some target hypothesis choice $h^* \in V(\hat{S})$. This contradicts the assumption there is a correct strategy with worst case cost C^* . \square

We now establish that when GCC is small, there must be a question which increases \bar{F}_α .

Lemma 3. *For any initial set of questions-response pairs \hat{S} , there must be a question $q \in Q$ such that*

$$\min_{h \in V(\hat{S})} \min_{r \in q(h)} \bar{F}_\alpha(\hat{S} + (q, r)) - \bar{F}_\alpha(\hat{S}) \geq c(q)(\alpha - \bar{F}_\alpha(\hat{S}))/GCC$$

Proof. Assume the lemma is false and for every question q there is some $h \in V(\hat{S})$ and $r \in q(h)$ such that

$$\bar{F}_\alpha(\hat{S} + (q, r)) - \bar{F}_\alpha(\hat{S}) < c(q)(\alpha - \bar{F}_\alpha(\hat{S}))/GCC$$

Define an oracle T' which answers every question with a response satisfying this inequality. For example, one such T' is

$$T'(q) \triangleq \operatorname{argmin}_r \bar{F}_\alpha(\hat{S} + (q, r)) - \bar{F}_\alpha(\hat{S})$$

By the definition of GCC

$$\min_{\hat{Q}: \bar{F}_\alpha(T'(\hat{Q})) \geq \alpha} c(\hat{Q}) \leq \max_{T \in R^Q} \left(\min_{\hat{Q}: \bar{F}_\alpha(T(\hat{Q})) \geq \alpha} c(\hat{Q}) \right) = GCC$$

so there must be a sequence of questions \hat{Q} with $c(\hat{Q}) \leq GCC$ such that $\bar{F}_\alpha(T'(\hat{Q})) \geq \alpha$. Because \bar{F}_α is monotone non-decreasing, we also know $\bar{F}_\alpha(T'(\hat{Q}) \cup \hat{S}) \geq \alpha$. Using the submodularity of \bar{F}_α ,

$$\begin{aligned} & \bar{F}_\alpha(T'(\hat{Q}) \cup \hat{S}) \\ & \leq \bar{F}_\alpha(\hat{S}) + \sum_{q \in \hat{Q}} (\bar{F}_\alpha(\hat{S} \cup \{(q, T(q))\}) - \bar{F}_\alpha(\hat{S})) \\ & < \bar{F}_\alpha(\hat{S}) + \sum_{q \in \hat{Q}} c(q)(\alpha - \bar{F}_\alpha(\hat{S}))/GCC \leq \alpha \end{aligned}$$

which is a contradiction. \square

We can now show approximate optimality.

Theorem 1. *Assume that α is an integer and, for any $h \in H$, F_h is an integral monotone non-decreasing submodular function. Algorithm 1 incurs at most $GCC(1 + \ln(\alpha n))$ cost.*

Proof. Let \hat{q}_i be the question asked on the i th iteration, \hat{S}_i be the set of question-response pairs after asking \hat{q}_i and C_i be $\sum_{j \leq i} c(\hat{q}_j)$. By Lemma 3

$$\bar{F}_\alpha(\hat{S}_i) - \bar{F}_\alpha(\hat{S}_{i-1}) \geq c(\hat{q}_i)(\alpha - \bar{F}_\alpha(\hat{S}_{i-1}))/GCC$$

After some algebra we get

$$\alpha - \bar{F}_\alpha(\hat{S}_i) \leq (\alpha - \bar{F}_\alpha(\hat{S}_{i-1}))(1 - c(\hat{q}_i)/GCC)$$

Now using $1 - x < e^{-x}$

$$\alpha - \bar{F}_\alpha(\hat{S}_i) \leq (\alpha - \bar{F}_\alpha(\hat{S}_{i-1}))e^{-c(\hat{q}_i)/GCC} = \alpha e^{-C_i/GCC}$$

We have shown that the gap $\alpha - \bar{F}_\alpha(\hat{S}_i)$ decreases exponentially fast with the cost of the questions asked. The remainder of the proof proceeds by showing that (1) we can decrease the gap to $1/|H|$ using questions with at most $GCC \ln(\alpha|H|)$ cost and (2) we can decrease the gap from $1/|H|$ to 0 with one question with cost at most GCC .

Let j is the largest integer such that $\alpha - \bar{F}_\alpha(\hat{S}_j) \geq 1/|H|$ holds. Then

$$1/|H| \leq \alpha e^{-C_j/GCC}$$

Solving for C_j we get $C_j \leq GCC \ln(\alpha|H|)$. This completes (1).

By Lemma 3, $\bar{F}_\alpha(\hat{S}_i) < \bar{F}_\alpha(\hat{S}_{i+1})$ (we strictly increase the objective on each iteration). Because α is an integer and for every h F_h is an integral function, we can conclude $\bar{F}_\alpha(\hat{S}_i) < \bar{F}_\alpha(\hat{S}_{i+1}) + 1/|H|$. Then q_{j+1} will be the final question asked. By Lemma 3, q_{j+1} can have cost no greater than GCC . This completes (2). We can finally conclude the cost incurred by the greedy algorithm is at most $GCC(1 + \ln(\alpha|H|))$ \square

By combining Theorem 1 and Lemma 2 we get

Corollary 1. *For integer α and integral monotone non-decreasing submodular F_h , the worst case cost of Algorithm 1 is within $1 + \ln(\alpha|H|)$ of that of any other correct question asking strategy*

We have shown a result for integer valued α and objective functions. We speculate that for more general non-integer objectives it should be possible to give results similar to those for standard submodular set cover (Wolsey, 1982). These approximation bounds typically add an additional normalization term.

6. Negative Results

The method we propose for interactive submodular set cover simultaneously solves the learning problem and covering problem in parallel, only solving the learning problem to the extent that it helps solve the covering problem. A simpler strategy is to solve these two problems in series (i.e. first identify h^* using the standard greedy query learning algorithm and second solve the submodular set cover problem for F_{h^*} using the standard greedy set cover algorithm). We call this the

Learn then Cover approach. We can show that this approach and in fact any approach that identifies h^* exactly can perform very poorly. Therefore it is important to consider the learning problem and covering problem simultaneously.

Theorem 2. *Assume F_h is integral for all h and that α is an integer. Any algorithm that exactly identifies h^* has approximation ratio at least $\Omega(|H| \max_i c(q_i) / \min_i c(q_i))$.*

Another simple approach is to ignore feedback and solve the covering problem for all $h \in H$. We call this the Cover All method. This method is an example of a non-adaptive method: a non-adaptive (i.e. non interactive) method is any method that does not use responses to previous questions in deciding which question to ask next. The *adaptivity gap* (Dean et al., 2004) for a problem characterizes how much worse the best non-adaptive method can perform as compared to the best adaptive method. For interactive submodular set cover we define the adaptivity gap to be the maximum ratio between the cost of the optimal non-adaptive strategy and the optimal adaptive strategy. With this definition, we can show that, in contrast to related problems (Asadpour et al., 2008) where the adaptivity gap is a constant, the adaptivity gap for interactive submodular set cover is quite large.

Theorem 3. *The adaptivity gap for interactive submodular set cover is at least $\Omega(|H| / \ln |H|)$.*

This result shows, even if we optimally solve the submodular set cover problem, the Cover All method can incur exponentially greater cost than the optimal adaptive strategy.

We can finally also show that the $1 + \ln(\alpha|H|)$ approximation factor achieved by the method we propose is in fact the best possible up to the constant factor assuming there are no slightly superpolynomial time algorithms for NP. The result and proof are very similar to those for the non-interactive setting (Krause et al., 2008).

Theorem 4. *Interactive submodular set cover cannot be approximated within a factor of $(1 - \epsilon) \max(\ln |H|, \ln \alpha)$ in polynomial time for any $\epsilon > 0$ unless NP has $n^{O(\log \log n)}$ time deterministic algorithms.*

The approximation factor we have shown for the greedy algorithm is

$$1 + \ln(\alpha|H|) = 1 + \ln \alpha + \ln |H| < 1 + 2 \max(\ln |H|, \ln \alpha)$$

so our hardness of approximation result matches up to the constant factor and lower order term.

7. Experiments

We tested our method on the interactive dominating set problem described in Section 4. In this problem, we are given a graph and H is a set of possibly overlapping clusters of nodes. The goal is to find a small set of nodes which forms a dominating set of an initially unknown target group $h^* \in H$. After selecting each node, we receive feedback indicating if the selected node is in the target group. Our proposed method (Simultaneous Learning and Covering) simultaneously learns about the target group h^* and finds a dominating set for it. We compare to two baselines: a method which first exactly identifies h^* and then finds a dominating set for the target group (Learn then Cover) and a method which simply ignores feedback and finds a dominating set for the union of all clusters (Cover All). Note that Theorem 2 and Theorem 3 apply to Learn then Cover and Cover All respectively, so these methods do not have strong theoretical guarantees. However, we might hope however that for reasonable real world problems they perform well. We use real world network data sets with simple synthetic hypothesis classes designed to illustrate differences between the methods. The networks are from Jure Leskovec’s collection of datasets available at <http://snap.stanford.edu/data/index.html>. We convert all the graphs into undirected graphs and remove self edges.

Table 1 shows our results. Each reported result is the average number of queries over 100 trials. Bolded results are the best methods for each setting with multiple results bolded when differences are not statistically significant (within $p = .01$ with a paired t-test). In the first set of results (Clusters), we create H by using the METIS graph partition package 4 separate times partitioning the graph into 10, 20, 30, and 40 clusters. H is the combined set of 100 clusters, and these clusters overlap since they are taken from 4 separate partitions of the graph. The target h^* is chosen at random from H . With this hypothesis class, we’ve found that there is very little difference between the Simultaneous Learning and Covering and the Learn then Cover methods. The Cover All method performs significantly worse because without the benefit of feedback it must find a dominating set of the entire graph.

In the second set of results, we use a hypothesis class designed to make learning difficult (Noisy Clusters). We start with H generated as before. We then add to H 100 additional hypotheses which are each very similar to h^* . Each of these hypotheses consists of the target group h^* with a random member removed. H is then the combined set of the 100 original hy-

Table 1. Average number of queries required to find a dominating set in the target group.

Data Set / Hypothesis Class	Simultaneous Learning and Covering	Learn then Cover	Cover All
Enron / Clusters	156.64	161.81	3091.00
Physics / Clusters	175.97	177.88	3340.00
Physics Theory / Clusters	172.38	175.12	3170.00
Epinions / Clusters	774.81	779.23	15777.00
Slashdot / Clusters	709.30	715.39	15383.00
Enron / Noisy Clusters	179.00	231.03	3091.00
Physics / Noisy Clusters	186.13	225.02	3340.00
Physics Theory / Noisy Clusters	160.62	201.24	3170.00
Epinions / Noisy Clusters	788.52	788.06	15777.00
Slashdot / Noisy Clusters	804.87	804.86	15383.00

hypotheses and these 100 variations of h^* . For this hypothesis class, Learn then Cover performs significantly worse than our Simultaneous Learning and Covering method on 3 of the 5 data sets. Learn then Cover exactly identifies h^* , which is difficult because of the many hypotheses similar to h^* . Our method learns about h^* but only to the extent that it is helpful for finding a small dominating set. On the other two data sets Learn then Cover and Simultaneous Learning and Covering are almost identical. These are larger data sets, and we’ve found that when the covering problem requires many more queries than the learning problem, our method is nearly identical to Learn then Cover. This makes sense since when α is large compared to the sum over $F_h(\hat{S})$ the second term in \bar{F}_α dominates.

Although we use real world graph data, the hypothesis classes and target hypotheses we use are very simple and synthetic, and as such these experiments are primarily meant to provide reasonable examples in support of our theoretical results.

8. Future Work

We believe there are other interesting applications which can be posed as interactive submodular set cover. In some applications it may be difficult to compute \bar{F}_α exactly because H may be very large or even infinite. In these cases, it may be possible to approximate this function by sampling from H . It’s also important to consider methods that can handle misspecified hypothesis classes and noise within the learning. One approach could be to extend agnostic active learning (Balcan et al., 2006) results to a similar interactive optimization setting.

References

Asadpour, A., Nazerzadeh, H., and Saberi, A. Stochastic submodular maximization. In *Workshop on Internet and Network Economics*, 2008.

Balcan, M.F., Beygelzimer, A., and Langford, J. Agnostic

active learning. In *ICML*, 2006.

- Balcázar, J.L., Castro, J., Guijarro, D., Köbler, J., and Lindner, W. A general dimension for query learning. *Journal of Computer and System Sciences*, 73(6):924–940, 2007.
- Dean, B., Goemans, M., and Vondrak, J. Approximating the stochastic knapsack problem: The benefit of adaptivity. In *FOCS*, 2004.
- Feige, U. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4), 1998.
- Goemans, M. and Vondrák, J. Stochastic covering and adaptivity. In *LATIN*, 2006.
- Golovin, D. and Krause, A. Adaptive submodularity: A new approach to active learning and stochastic optimization. In *COLT*, 2010.
- Guillory, Andrew and Bilmes, Jeff. Interactive submodular set cover, 2010. Technical Report. UWEETR-2010-0001.
- Hanneke, S. The cost complexity of interactive learning, 2006. Unpublished. <http://www.stat.cmu.edu/~shanneke/docs/2006/cost-complexity-working-notes.pdf>.
- Kempe, D., Kleinberg, J., and Tardos, É. Maximizing the spread of influence through a social network. In *KDD*, 2003.
- Kempe, D., Kleinberg, J., and Tardos, É. Influential nodes in a diffusion model for social networks. In *ICALP*, 2005.
- Krause, A., McMahan, H.B., Guestrin, C., and Gupta, A. Robust submodular observation selection. *JMLR*, 2008.
- Lin, Hui and Bilmes, Jeff. Multi-document summarization via budgeted maximization of submodular functions. In *NAACL/HLT*, 2010.
- Narayanan, H. *Submodular Functions and Electrical Networks*. North Holland, 1997.
- Wolsey, L.A. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2(4), 1982.