Simultaneous Learning and Covering with Adversarial Noise

Andrew Guillory
Department of Computer Science
University of Washington
guillory@cs.washington.edu

Jeff Bilmes
Department of Electrical Engineering
University of Washington
bilmes@ee.washington.edu

Abstract

We study simultaneous learning and covering problems which are set cover problems that depend on the solution to a learning problem. The goal is to jointly minimize the cost of both learning and covering. We extend previous work to allow for a limited amount of adversarial noise. As a special case our problem can be applied to noisy query learning.

1 Background

Consider a social network advertising problem where we want to send ads to a small set of users that are influential in a target group (e.g. users that like jazz music). If we knew the members of the target group we could solve this problem as a submodular function maximization problem [7] (i.e., maximizing the influence of a fixed size set of users) or a submodular set cover problem (i.e., minimizing the size of a fixed influence set). Other applications like document summarization [9] and sensor placement [8] can also be solved via submodular function maximization or submodular set cover.

In this work we assume instead that we do not initially know which users are in the target group. Then we must both identify the target group (*learn*) and send ads to a small set of users influential in that group (*cover*). Our goal is to minimize the joint cost of both learning and covering. For example, a simple interaction model could assume that after sending a user an ad we are told in the form of ad feedback (e.g. ad clicks) whether or not that user was a member of the group of interest (e.g. interested in jazz). In this case learning actions and covering actions are one and the same.

We can also consider variations of this problem in which there are separate learning actions corresponding to querying users (asking a user "Do you like jazz?") and advertising actions which may or may not reveal information about the target group. Other applications outside of advertisement can also be phrased in terms of simultaneous learning and covering. For example we may want to find a set of documents that summarizes all documents that a user is interested in; if we do not initially know which documents a user is interested in then we must both identify the documents of interest (learn) and also summarize them (cover).

In previous work [5], we introduced and analyzed a problem called *interactive submodular set cover*, a direct generalization of both exact active learning with a finite hypothesis class (query learning) and submodular set cover. Many simultaneous learning and covering type problems can be formally posed as interactive submodular set cover problems. Interactive submodular set cover can be seen as a unification of previous work on approximation algorithms for submodular set cover [10] and approximation algorithms for query learning [6, 1]. We formally state interactive submodular set cover below.

Interactive Submodular Set Cover
Given:

- Hypothesis class H containing unknown hypothesis h^*
- Query set Q and response set R with known $q(h) \subseteq R$ for $q \in Q$, $h \in H$
- Modular query cost function c defined over Q
- Submodular, monotone objective functions $F_h : Q \times R \to \mathbb{R}_{\geq 0}$ for $h \in H$
- 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 $\sum_i c(\hat{q}_i)$

In the advertisement example described above, the hypothesis class H corresponds to all possible target groups, and the target hypothesis $h^* \in H$ is the (initially unknown) target group. Queries correspond to actions available to the learning and covering algorithm; for example, we can have an action for each user corresponding to sending that user an ad. Responses correspond to feedback from actions, and $q(h) \subseteq R$ is the set of allowable responses for question q and hypothesis h. For example, an advertising action with feedback could return 1 if the user is in the target group and 0 otherwise. We assume $q(h) \neq \emptyset$ and that $q(h)$ is known (and available to an algorithm) for every q and h.

Finally, the objective function $F_h(\hat{S})$ for hypothesis h measures the influence achieved by the set of query-response pairs (e.g. advertising actions) $\hat{S}$ if hypothetically h were the target hypothesis (target group). We assume F_h is submodular and monotone non-decreasing and known. A set function $F : 2^V \to \mathbb{R}$ is submodular iff for every $A \subseteq B \subseteq (V - s)$

$$F(A + s) - F(A) \geq F(B + s) - F(B)$$

In other words, the gain of performing an action decreases as we perform other actions (F exhibits diminishing returns).

The goal is to ensure $F_{h^*}(\hat{S}) \geq \alpha$ for a fixed α (a target level of influence is achieved in the target group) using minimal cost . Note that we assume that F_h is known for every h; what we do not know is h^*, which of the hypotheses is the target hypothesis. Define the *version space* $V(\hat{S})$ to be the subset of H consistent with all query-response pairs in $\hat{S}$. More formally,

$$V(\hat{S}) \triangleq \{h \in H : \forall (q, r) \in \hat{S},\ r \in q(h)\}$$

For worst case choice of h^*, in order to ensure that $F_{h^*}(\hat{S}) \geq \alpha$ it is both necessary and sufficient to ensure that $F_h(\hat{S}) \geq \alpha$ for every $h \in V(\hat{S})$.

Interactive submodular set cover makes the limiting, simplifying assumption that the target hypothesis h^* is in H. In terms of the advertising example, we assume that the community of interest is one of a finite set of communities we know in advance and that there is no noise in our feedback. In almost every real world learning application, this is an unrealistic assumption; typically, no single hypothesis will exactly match up with the observed question-response pairs because of noise or because we have an incorrect hypothesis class.

2 Noisy Interactive Set Cover

In this work we propose a generalization of interactive submodular set cover that relaxes the assumption that $h^* \in H$. With h^* not necessarily in H, it no longer makes sense to require $F_{h^*}(\hat{S}) \geq \alpha$; we only know objective functions F_h for $h \in H$ so if $h^* \notin H$ then we have no way of testing if $F_{h^*}(\hat{S}) \geq \alpha$. Instead we require that the covering constraint $F_h(\hat{S}) \geq \alpha$ is satisfied for all h that are sufficiently "close" to h^*.

We propose defining this closeness in terms of an additional submodular, monotone non-decreasing function $G_h(\hat{S})$ defined over question-response pairs. In particular, we require that the covering constraint $F_h(\hat{S}) \geq \alpha$ is satisfied for all h such that $G_h(S^*) < \kappa$ where κ is a known constant and

S^* is the unknown set of all question-response pairs induced by h^*, $S^* \triangleq \bigcup_{q \in Q, r \in q(h^*)} \{(q,r)\}$. Intuitively, G_h can be thought of as computing the distance from h to h^* in terms of the question-response pairs allowed by h^*. We make the assumption that for any $\hat{S}$, and (q,r) such that $r \in q(h)$, $G_h(\hat{S}+(q,r)) - G_h(\hat{S}) = 0$. In other words, if we observe a question-response pair that agrees with h, then G_h does not increase. The constant κ is the threshold which determines which hypotheses are sufficiently close.

Noisy Interactive Set Cover

Given:

- Hypothesis class H (does not necessarily contain h^*)
- Query set Q and response set R with known $q(h) \subseteq R$ for $q \in Q$, $h \in H$ and unknown $q(h^*) \subseteq R$ for $q \in Q$
- Modular query cost function c defined over Q
- Submodular, monotone, normalized $F_h : Q \times R \to \mathbb{R}_{\geq 0}$ for $h \in H$
- Submodular, monotone, normalized $G_h : Q \times R \to \mathbb{R}_{\geq 0}$ for $h \in H$ with

$$G_h(S + (q,r)) - G_h(S) = 0$$

  for any S if $r \in q(h)$
- Objective threshold α, closeness threshold κ

Repeat: Ask a question $\hat{q}_i \in Q$ and receive a response $\hat{r}_i \in \hat{q}_i(h^*)$
Until: $F_h(\hat{S}) \geq \alpha$ for all h such that $G_h(S^*) < \kappa$ where $\hat{S} = \bigcup_i \{(\hat{q}_i, \hat{r}_i)\}$ and $S^* \triangleq \bigcup_{q \in Q, r \in q(h^*)} \{(q,r)\}$
Objective: Minimize $\sum_i c(\hat{q}_i)$

By setting $\kappa = 1$ and using $G_h(\hat{S}) \triangleq I(h \notin V(\hat{S}))$ we recover a variation of the original interactive submodular set cover problem. In this case satisfying the covering constraint for all h that agree exactly with h^* corresponds exactly to satisfying the covering constraint for all h such that $G_h(S^*) < \kappa$. This is equivalent to the interactive submodular set cover problem if we include the optional assumption $G_h(S^*) < \kappa$ for at least one h (i.e. h^* agrees with some $h \in H$). We will in fact argue that the algorithm we propose is approximately optimal regardless of whether or not this additional assumption is made.

We can interpret this more general problem as using an extended notion of version space such that a hypothesis h is no longer immediately eliminated as soon as a question-response pair $(\hat{q}_i, \hat{r}_i)$ with $\hat{r}_i \notin \hat{q}_i(h)$ is observed. Instead, a hypothesis h is only eliminated from consideration when $G_h(\hat{S}) \geq \kappa$. Different definitions of κ and G_h correspond to different notions of version space. This intuition is formalized by the following lemma.

Lemma 1. *For worst case choice of h^*, to ensure the covering constraint $F_h(\hat{S}) \geq \alpha$ is satisfied for all h such that $G_h(S^*) < \kappa$ it is both necessary and sufficient to ensure that the covering constraint $F_h(\hat{S}) \geq \alpha$ is satisfied for all h such that $G_h(\hat{S}) < \kappa$*

Proof. To see this condition is sufficient note that $G_h(\hat{S}) \leq G_h(S^*)$. To see this condition is necessary note that for any particular h this inequality holds with equality for some h^* (the h^* which agrees with h on queries not in $\hat{Q}$). We cannot therefore eliminate any h for which $G_h(\hat{S}) < \kappa$ as for some choice of h^* this hypothesis also satisfies $G_h(S^*) < \kappa$. □

3 Application to Noisy Query Learning

Like interactive submodular set cover, noisy interactive set cover is related to query learning. In particular, we recover a version of query learning by using

$$F_h(\hat{S}) \triangleq |H \setminus V(\hat{S})| \quad \text{and} \quad G_h(\hat{S}) \triangleq I(h \notin V(\hat{S}))$$

with $\kappa = 1$ and $\alpha = |H| - 1$. For these objectives, the goal of noisy interactive set cover is to eliminate $|H| - 1$ hypotheses from the version space. This is equivalent to standard query learning if we make the additional assumptions that (1) there is an $h \in H$ that agrees with h^* on every query and (2) for each $h, h' \in H$ there is some $q \in Q$ with $q(h) \cap q(h') = \emptyset$. With these two assumptions, it is always possible to eliminate $|H| - 1$ hypotheses from the version space and the remaining hypothesis is the one hypothesis that agrees with h^*.

Noisy interactive set cover can also be used to extend query learning to allow for noise. Define $\mathrm{err}(h, \hat{S}) \triangleq \sum_{(q,r)\in\hat{S}} I(r \notin q(h))$. By using

$$F_h(\hat{S}) \triangleq \sum_{h'\neq h} \min(\mathrm{err}(h', \hat{S}), \kappa) \quad \text{and} \quad G_h(\hat{S}) \triangleq \mathrm{err}(h, \hat{S})$$

and $\alpha = \kappa(|H| - 1)$ the goal of noisy interactive set cover is to identify $|H| - 1$ hypotheses that make at least κ mistakes with respect to h^*. We can then assume (1) that at least one h satisfies $G_h(S^*) < \kappa$ (makes less than κ mistakes with respect to h^*) and (2) every pair of hypotheses $h, h' \in H$ have $q(h) \cap q(h') = \emptyset$ for at least $2(\kappa - 1) + 1$ different $q \in Q$. With these assumptions, it is always possible to identify $|H| - 1$ hypotheses that make at least κ mistakes and the remaining h must make less than κ mistakes. Then noisy interactive set cover solves a problem where the goal is to identify the one $h \in H$ that makes fewer than κ errors.

Dasgupta et al. [2] study a similar query learning setting with adversarial noise–the setting they consider is the same except that they do not require that pairs of hypotheses in H are distant and instead simply ask that the learning algorithm return any h with $G_h(S^*) < \kappa$ when multiple exist. When applied to noisy query learning, our approximation guarantees are similar to theirs (we get an $O(\ln(\kappa|H|))$ approximation ratio while they get $O(\ln |H|)$–we suspect our dependence on κ is due to allowing G_h to be submodular). Recently Golovin et al. [4] have extended their previous work on adaptive submodularity [3] to give approximation results for an average-case noisy query learning setting (the setting we consider here is worst case).

We note that we will show our algorithm is approximately optimal whether or not we assume (1) (restricting the worst case h^* in this way does not hurt the approximation). Also, assuming (2) has no effect on the analysis since it is an assumption about the hypothesis class H and our analysis holds for any H. As an alternative to making these additional assumptions we could instead modify the objective function F_h to exclude from the sum h' that are "close" to h. With this variation the goal is to identify a subset of H such that all pairs of h, h' in the subset are close to each other and all h not in the set make at least κ mistakes. This is more similar to the setting of Dasgupta et al. [2].

4 Greedy Algorithm and Analysis

Guillory and Bilmes [5] showed that a simple greedy algorithm is approximately optimal for interactive submodular set cover. The analysis proceeds in two steps. First, the optimization problem over many objectives F_h is reduced to a simpler problem over a single objective $\bar{F}_\alpha$. Second, a greedy algorithm for this simpler problem is proposed and analyzed. In this section we show that a similar analysis holds for noisy interactive set cover. The key insight is that noisy interactive set cover can also be reduced to a problem over a (different) single objective $\bar{F}_\alpha$ which is submodular and monotone non-decreasing.

Define the following objective

$$\bar{F}_\alpha(\hat{S}) \triangleq \frac{1}{\kappa|H|} \sum_{h\in H} \Big((\kappa - \min(G_h(\hat{S}), \kappa)) \min(F_h(\hat{S}), \alpha) + \min(G_h(\hat{S}), \kappa)\alpha \Big)$$

$\bar{F}_\alpha(\hat{S}) \geq \alpha$ iff this set of query response pairs solves the noisy interactive set cover problem. That is, $\bar{F}_\alpha(\hat{S}) \geq \alpha$ iff $F_h(\hat{S}) \geq \alpha$ for all h such that $G_h(\hat{S}) < \kappa$. From Lemma 1 this is sufficient and necessary to ensure $F_h(\hat{S}) \geq \alpha$ for all h such that $G_h(S^*) < \kappa$ for worst case h^*.

We can then define a greedy algorithm for solving the simplified problem over $\bar{F}_\alpha$. See Algorithm 1. At each time step the algorithm performs the query $q_i \in Q$ that maximizes the worst case normalized gain

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

This algorithm is the same as the worst case greedy algorithm for interactive submodular set cover with the exceptions that the objective function $\bar{F}_\alpha$ is different and the response for a question q_i is

Algorithm 1 Worst Case Greedy

$\hat{S} \Leftarrow \emptyset$
while $\bar{F}_\alpha(\hat{S}) < \alpha$ **do**
  $\hat{q} \Leftarrow \text{argmax}_{q_i \in Q} \min_{r_i \in R}(\bar{F}_\alpha(\hat{S} + (q_i, r_i)) - \bar{F}_\alpha(\hat{S}))/c(q_i)$
  Ask $\hat{q}$ and receive response $\hat{r}$
  $\hat{S} \Leftarrow \hat{S} + (\hat{q}, \hat{r})$
end while

no longer required to match the response set $q_i(h)$ for some $h \in H$ (responses are arbitrary since h^* may not be in H). Crucial to the analysis is the fact that $\bar{F}_\alpha(S)$ is submodular.

Lemma 2. *Let $F(S)$ be a monotone non-decreasing submodular function ranging between* 0 *and α. Let $G(S)$ be a monotone non-decreasing submodular function ranging between* 0 *and κ. Then*

$$\bar{F}(S) \triangleq (\kappa - G(S))F(S) + G(S)\alpha$$

is a monotone non-decreasing submodular function.

Proof. As a short hand we use $\text{Gain}(F, S, x) \triangleq F(S + x) - F(S)$. We show $\bar{F}(S)$ is monotone non-decreasing by showing that for any A and x, $\text{Gain}(\bar{F}, A, x) \geq 0$.

$$\begin{aligned}
&\text{Gain}(\bar{F}, A, x) \\
&\quad = \quad (\kappa - G(A + x))F(A + x) + G(A + x)\alpha - (\kappa - G(A))F(A) - G(A)\alpha \\
&\quad = \quad \kappa \,\text{Gain}(F, A, x) + G(A)F(A) - G(A + x)F(A + x) + \text{Gain}(G, A, x)\alpha \\
&\quad = \quad (\kappa - G(A))\,\text{Gain}(F, A, x) + \text{Gain}(G, A, x)(\alpha - F(A + x))
\end{aligned}$$

We see that each of these terms is positive so long as F and G are monotone non-decreasing and range between 0 and α and 0 and κ respectively. We now show $\bar{F}(S)$ is submodular by showing, for any $A \subseteq B$ and x, $\text{Gain}(\bar{F}, A, x) \geq \text{Gain}(\bar{F}, B, x)$. As before we have

$$\text{Gain}(\bar{F}, B, x) = (\kappa - G(B))\,\text{Gain}(F, B, x) + \text{Gain}(G, B, x)(\alpha - F(B + x))$$

Each term in this expression is less than the corresponding term in $\text{Gain}(\bar{F}, A, x)$. □

Corollary 1. *$\bar{F}_\alpha(S)$ is submodular monotone non-decreasing whenever F_h and G_h are submodular monotone non-decreasing for all $h \in H$.*

Proof. The result follows from Lemma 2 and the following known results. (1) When a function $F(S)$ is submodular and monotone non-decreasing, so is $F_\alpha(S) = \min(F(S), \alpha)$ for a constant α. (2) When $F_h(S)$ is submodular and monotone non-decreasing for every $h \in H$, $\bar{F}(S) = \sum_{h \in H} F_h(S)$ is also submodular and monotone non-decreasing. □

Having established the submodularity of $\bar{F}_\alpha$, the remainder of the analysis mostly follows that for interactive submodular set cover. The one difference is that the responses we receive may be arbitrary since we no longer assume $h^* \in H$. However, we will in fact show that our proposed algorithm is approximately optimal whether or not we assume $G_h(S^*) < \kappa$ for some h (assume that h^* must be close to some $h \in H$). Recall this assumption was useful when posing noisy query learning as noisy interactive set cover. Our analysis is based on the Extended Teaching Dimension analysis of query learning [6, 1].

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 [6]. Define the General Cover Cost, GCC

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

GCC depends on H, Q, α, κ, c, and the objective functions F_h and G_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* $\hat{Q}$.

We first show that GCC is a lower bound on the optimal worst case cost of satisfying $\bar{F}_\alpha(\hat{S}) \geq \alpha$. We've already argued this is necessary and sufficient for solving the noisy interactive set cover problem.

Lemma 3. *If there is a correct question asking strategy for satisfying $\bar{F}_\alpha(\hat{S}) \geq \alpha$ 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 $\bar{F}_\alpha(\hat{S}) < \alpha$. This contradicts the assumption there is a correct strategy with worst case cost C^*. □

We show that in fact $GCC \leq C^*$ even if we make the additional assumption that some for some h $G_h(S^*) < \kappa$. This is a stronger result since C^* is smaller in this case as we have placed a restriction on h^*.

Lemma 4. *Assume for some h $G_h(S^*) < \kappa$. If there is a correct question asking strategy for satisfying $\bar{F}_\alpha(\hat{S}) \geq \alpha$ with worst case cost C^* then $GCC \leq C^*$.*

Proof. Assume as before the lemma is false and there is a correct question asking strategy with worst case cost C^* and $GCC > C^*$. As before there is a T^* such that any sequence of questions $\hat{Q}$ with total cost less than or equal to C^* must have $\bar{F}_\alpha(\hat{S}) < \alpha$. Then, there must be some remaining h with $G_h(\hat{S}) < \kappa$ and $F_h(\hat{S}) < \alpha$. Therefore, for any sequence of questions with cost less than or equal to C^*, there is some h^* such that $G_h(S^*) < \kappa$ for some h and it is possible to answer questions consistently with h^* such that $G_h(\hat{S}) < \kappa$ and $F_h(\hat{S}) < \alpha$. In particular, any h^* with $q(h^*) = \{T^*(q)\}$ for $q \in \hat{Q}$ and $q(h^*) = q(h)$ for $q \notin \hat{Q}$ suffices. □

The rest of the analysis follows that for interactive submodular set cover [5] and gives the following.

Theorem 1. *For integer α and κ and integral monotone, normalized, submodular F_h and G_h, the worst case cost of Algorithm 1 is within $1 + \ln(\kappa\alpha|H|)$ of that of any other correct question asking strategy. This holds whether or not we assume $G_h(S^*) < \kappa$ for some h.*

References

[1] J. Balcázar, J. Castro, D. Guijarro, J. Köbler, and W. Lindner. A general dimension for query learning. *Journal of Computer and System Sciences*, 73(6):924–940, 2007.

[2] S. Dasgupta, W. Lee, and P. Long. A theoretical analysis of query selection for collaborative filtering. *Machine Learning*, 51(3), 2003.

[3] D. Golovin and A. Krause. Adaptive submodularity: A new approach to active learning and stochastic optimization. In *COLT*, 2010.

[4] D. Golovin, A. Krause, and D. Ray. Near-optimal bayesian active learning with noisy observations. In *NIPS*, 2010.

[5] A. Guillory and J. Bilmes. Interactive submodular set cover. In *ICML*, 2010.

[6] S. Hanneke. The cost complexity of interactive learning, 2006. Unpublished. `http://www.stat.cmu.edu/~shanneke/docs/2006/cost-complexity-working-notes.pdf`.

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

[8] A. Krause, H. McMahan, C. Guestrin, and A. Gupta. Robust submodular observation selection. *JMLR*, 2008.

[9] H. Lin and J. Bilmes. Multi-document summarization via budgeted maximization of submodular functions. In *NAACL/HLT*, 2010.

[10] L. Wolsey. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2(4), 1982.