Ensemble Interactive Submodular Set Cover in Regression Scenarios

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Submodular set cover problem has been well studied in many different scenarios. A interactive variant of it take in the virtue of active learning and enabled us to gain satisfitory submodular utilities of all sufficiently plausible hypotheses with a lowest cost of actions. Moreover, in the smooth extension of interactive submodular set cover introduced a more flexible algorithm under settings where the covering goal varies smoothly under agnostic circumstance and bring light to approximately optimize over real-valued submodular functions. In this paper, we propose an iterative ensemble method to deal with poor prior knowledge (i.e., poor base hypotheses) which should influence performance of covering in previous work. With a little reasonable assuming we naturally extend the submodular set cover problem to the regression scenario, which means the covering goal is no longer including elements in a particular class but points on a curve sharing similar criteria of utilities and also provide a solution to the infinite hypothesis class problem with additional linear constraints common in regression scenarios.

1 Introduction

General interactive submodular set cover [6] [4] aims to interactively distinguish a true submodular hypothesis from a finite hypothesis set and satisfy the utility of it to a target threshold with as low cost as possible. Under an agnostic setting [5], which relaxes the assumption that the hypothesis set necessarily contains a true hypothesis, and turn to satisfy all plausible hypotheses. Recently, In [7], the algorithm became more flexible by considering all hypotheses and the submodular utility function has been generalized to a real-valued one.

An example of interactive submodular set cover is in viral marketing, sending advertisments to a minimal set of people to achieve a desired influence. The advertiser does not has the priori knowledge of whether a particular customer is interested in the advertisments but can interactively learn it after they send advertisments and receive the feedback. Hence the nature of interactive submodular set cover problems is to find a best decision sequence to trade off between exploration and exploitation, where exploration means to reduce the uncertainty regarding the true hypothesis and exploitation refers to achieve a relatively high utility.

In this paper we consider the problem of using a ensemble method to generate a better hypothesis than base ones iteratively and make selection based on this ensemble hypothesis in each step. In previous work, the choices are all based on objective functions unifying utility functions and distance functions of base hypotheses, and this ask for a good priori knowledge to attain high utility, otherwise algorithm usually make poor choice and stop when the disagreements are large and utilities are small. Our solution is to use a boosting-like method, perform non-negative linear combination over base hypotheses and update the ensemble parameters every time a new query-response pair is added.

Submitted to 29th Conference on Neural Information Processing Systems (NIPS 2016). Do not distribute.

We generally consider the regression-based setting, where the underlying exploration purpose is a regression problem rather than a classification one in previous work, and the utility of exploitation is no longer decided by the class a query point belongs to, but some continuously varied properties (such as slope of the curve at the very point). Under this enlarged setting, we need additional assumings to make the problem tractable. As a real-life application of such submodular set cover in regression, in AIDS prevention and cure, where the concerntration of CD4 varys over time and is a vital criterion to judge the effect of medical treatment. There was a proposal made by WHO that the demand for CD4 testing will grow strongly[1], so the CD4 concerntration can provide guidance to achieve a sufficently high curative effect within a limited medical resources, where the curative effect shows a nature of diminishing returns as times of treatment increase, i.e, submodularity.

We also explored the problem of infinite hypothesis class which is naturally arisen in regression scenarios, but it is overly complicated when the exact form of objective function is not defined. We provide a particular case of it with linear constraints.

To the best of our knowledge, it is the first time that using an ensemble method in interactive submodular set cover problems to combine individual hypothesis into a better one stepwise under a regression setting and extending the hypothesis space to a infinite one with constraints.

2 Related Work

Problems related to submodularity has been studied across many different goals and settings. optimization of submodular function can be generally devided into minimization and maximization. The unconstrianted submodular function minimization is possible to be solved in (strong) polynomial time [18] and there are a lot of work on it (c.f., [3]). Unlike the minimization problem, the maximization of submodular function [12] is usually NP-hard. Typically, the approximation algorithms for these problems are based on either greedy algorithms or local search algorithms. [16] proves that the greedy algorithm provides a good approximation to the optimal solution.

Submodular function optimization has shown promise for wide variety of applications and already been utilized in viral marketing [10], information gathering [13], image segmentation([2], [11], [9]), document summarization([14]), speeding up satisfiability solvers([19]) and many other domains.

Submodular set cover is closely related to submodular function maximization problem under a modular cost constraint. An excellent analysis of non-interactive submodular set cover problem and corresponding greedy algorithm was stated in [20].

Interactive variants of submodular set cover([4], [5], [7]) combine the basic problem with characteristic of active learning, which makes response returns to the decision maker immediately after each query asked. Meanwhile, presented in [6], many active learning problems canbe transferred into interactive submodular set cover problems.

Moreover, under regression settings, [1] proposed an efficient submodular function maximization algorithm in a streaming scenario and apply it to active subset selection, which helps in speeding up large-scale nonparametric regression problems.

3 Background

3.1 Submodular Functions

We firstly enumerate the definition of submodularity together with some other useful properties of submodular functions that will be used in Section 4
Give a finite *ground set* V and a set function $f : 2^V \to \mathbb{R}$ that assign each subset $S \subseteq V$ a value $f(S)$:

- The function f is *submodular* if for arbitrary $A \subseteq B \subseteq V$:

$$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

- The function f is *monotone non-decreasing* if $f(A) \leq f(B)$ for $A \subseteq B \subseteq V$.

[1]For more information, view `http://www.who.int/hiv/pub/posters/ias2015-poster2/en/`

- The function f is called *modular* (or *additive*, *linear*) if $f(A \cup \{e\}) = f(A) + f(\{e\})$ for $A \subseteq (V \setminus \{e\})$.
- The function f is *normalized* if $f(\emptyset) = 0$, and we always assume that f is normalized in the following sections.

3.2 Submodular Set Cover and its Interactive Variants

General submodular set cover problem [20] ask for satisfying a submodular utility function to a target threshold with a minimal cost of actions. Specifically, we are given a set of all actions Q (with each action $q \in Q$ associated with a modular cost function $c(q)$) and a monotone submodular utility function F. The goal of submodular set cover problem is to select a subset S of Q minimizing the cost function $c(S) = \sum_{s \in S} c(s)$ subject to the constraint $F(S) \geq \alpha$, where α is the target threshold metioned before.
This problem, sharing similar properties with the problem of submodular function maximization [12], is unfortunately NP-hard. Celebrately, a practically efficient greedy algorithem can approximately address this problem when F is integer-valued, and is able to achieve near-optimal cost of at most $(1 + \ln(\max_{e \in Q} F(\{e\})))OPT$.
In its interactive variant (named *Interactive Submodular Set Cover*, or ISSC for short), action-utility pairs in non-interactive scenarios are often called query-response pairs due to the utility of an action keeps unknown until the action is taken and a response from an oracle is received. In ISSC problems, a finite hypothesis class H containing an unknown true hypothesis h^* is given. So is a query set Q and a response set R, then each $q \in Q$ is a function mapping hypothesis class H to responeses R (i.e., $q(h) \in R$). An oracle is a response provider with each response $r \equiv q(h^*)$. Let S denote the set of observed query-response pairs: $S = \{(q, r)\} \subseteq Q \times R$, the basic process of ISSC is asking a q and receive an r from the oracle, then this iterates until $F_{h^*}(S) \geq \alpha$. The goal is to minimize the cost $c(S) = \sum_{(q,r) \in S} c(q)$. The ISSC problem is a simultaneously learning and covering prolem, which means one needs to balance between distinguishing h^* from other hypotheses and gaining a satisfactory utility.

More common in real life, in noisy ISSC [5], h^* is no longer assumed to be necessarily contained in H, and uses a distance function G_h and tolerance κ such that the goal is to satisfy $F_h(S) \geq \alpha$ for all sufficiently plausible h, where plausibility is defined as $G_h(S) \leq \kappa$. In a recent work, an extension named smooth ISSC [7] extend the noisy ISSC to enable smooth variation of the target threshold of the candidate submodular functions according to their plausibility and also propose a approximate version when the objective function is real-valued.

4 Ensemble Interactive Submodular Set Cover

4.1 Problem Statement and Notation

We consider the problem of iteratively obtaining an hypothesis $\hat{h}_S$ as accurate as possible to improve the performance in noisy situation (i.e., h^* is not necessarily included in H) based on the query/response pairs observed by applying a ensemble method to base hypotheses $h \in H$.
In this problem, some settings are generalized from noisy and smooth ISSC. Aside from the base hypothesis class H, we construct a augmented hypothesis class $H^+ \triangleq \{h \mid h = \sum_{i=1}^{|H|} \beta_i h_i + b, h_i \in H, \boldsymbol{\beta} \in \mathbb{R}_{\geq 0}^{|H|}, b \in \mathbb{R}\}$ that consists of all possible ensemble results of base hypothesis. Each $h \in H^+$ having two corresponding monotone non-decreasing submodular fuctions $F_h : Q \times R \rightarrow \mathbb{R}_{\geq 0}$ and $G_h : Q \times R \rightarrow \mathbb{R}_{\geq 0}$, where F_h represents total utility of query-response pairs observed and G_h evaluates closeness of particular h to h^*. Similarily, other concepts defined over H need to be extended to H^+. A query class Q and a response class R with known $q(h) \subseteq R$ for $q \in Q, h \in H^+$ is also given. For the threshold, We also require an exact threshold α for $F_{\hat{h}_{\hat{S}}}(\hat{S})$ to satisfy.

In *Ensemble Interactive Submodular Set Cover*, Let $S^* \triangleq \bigcup_{q \in Q, r \in q(h^*)} \{(q, r)\}$ denote the set of all possible query-response pairs given by h^*. The goal is to construct a query-response set $\hat{S}$ with minimal cost such that, for currently optimized ensemble hypothesis $\hat{h}_{\hat{S}}$ we have $F_{\hat{h}_{\hat{S}}}(\hat{S}) \geq \alpha$, where α is a given threshold.We summarize the general problem setting in Problem 4.1 and notation used in Table 1.

Table 1: Summary of notation

Variable	Definition
H	Set of hypotheses
H^+	Augmented set of hypotheses
Q	Set of queries
R	Set of responses
F_h	Monotone non-decreasing submodular utility function
G_h	Monotone non-decreasing submodular distance function
$\hat{h}_{\hat{S}}$	The estimated hypothesis after applying ensemble over the set $\hat{S}$
S_t	The set of observed query-response pairs after i-th iteration
S_T	The set of observed query-response pairs after algorithm terminated
α	Threshold that $F_{\hat{h}_{S_T}}(S_T)$ needs to satisfy

Problem 4.1. Ensemble Interactive Submodular Set Cover

Given:

1. Hypothesis class H (does not necessarily contain h^*) and augmented hypothesis class H^+
2. Query set Q and response set R with known $q(h) \subseteq R$ for $q \in Q, h \in H^+$
3. Modular query cost function c defined over Q
4. Monotone non-decreasing submodular objective functions $F_h : 2^{Q \times R} \to \mathbb{R}_{\geq 0}$ for $h \in H^+$
5. Monotone non-decreasing submodular distance functions $G_h : 2^{Q \times R} \to \mathbb{R}_{\geq 0}$ for $h \in H^+$, with $G_h(S \cup (q, r)) - G_h(S) = 0$ for any S if $r \in q(h)$

Protocol: For $i = 1, \ldots, |Q|$: ask a question $\hat{q}_i \in Q$ and receive a response $\hat{r}_i \in \hat{q}_i(h^*)$
Goal: Using minimal cost $\sum_i c(\hat{q}_i)$, terminate when $F_{\hat{h}_{\hat{S}}}(\hat{S}) \geq \alpha$, where $\hat{S} = \{(\hat{q}_i, \hat{r}_i)\}_i$, $S^* \triangleq \bigcup_{q \in Q, r \in q(h^*)} \{(q, r)\}$, α is a real-valued threshold below $\tau(G_{\hat{h}_{\hat{S}}}(S^*))$ and $\hat{h} = \operatorname*{arg\,min}_{h \in H^+} G_h(\hat{S})$

We prove a cost guarantees based on general cover cost(GCC) defined in Definition 4.2, For clarity of exposition, all proofs are deferred to the supplementary material.

Definition 4.2. (General Cover Cost (*GCC*)). *Define oracles $T \in R^Q$ to be functions mapping queries to responses and $T(\hat{Q}) \triangleq \bigcup_{\hat{q}_i \in \hat{Q}} \{(\hat{q}_i, T(\hat{q}_i))\}$. $T(\hat{Q})$ is the set of query-response pairs given by T for the set of queries $\hat{Q}$. Define the General Cover Cost as:*

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

Lemma 4.3. (Lemma 3 from [5]). *if there is a query asking strategy for satisfying $F_{\hat{h}_{\hat{S}}}(\hat{S}) \geq \alpha$ with worst case cost C^*, then $GCC \leq C^*$. Thus $GCC \leq OPT$.*

Lemma 4.3 obviously shows that GCC lower bounds the optimal cost.

4.2 Ensemble Interactive Submodular Set Cover in Finite Hypothesis Class

We firstly consider the problem of maximizing an objective function generated by applying an ensemble method over a finite hypothesis class version.

The exact form of the objective function is simply given as

Definition 4.4. (Exact Form of Objective Fuction).

$$F_{\hat{h}_S}(S) \triangleq \operatorname*{arg\,min}_{h \in H^+} G_h(S)$$

But a problem immediately arise that $F_{\hat{h}_{\hat{S}}}(\hat{S})$ is ever-changing along with each step we select a query and enlarge $\hat{S}$. We need to guarantee that for any fixed order of query choice, $F(S) = F_{\hat{h}_{\hat{S}}(\hat{S})}$ performs monotone non-decreasing submodularity and then this problem can be treated as an ISSC problem. This problem remains intractable if no constraints added to the updating of $\hat{h}_S$. We sacrific a little accuracy of $\hat{h}_S$ to make it the lowest disagreement-rate one in hypotheses that enable this problem be resolvable.

To exactly define this sacrific (or constraints), we firstly define a *Hypothesis Update Gain* as:

Definition 4.5. (Hypothesis Update Gain).

$$\Delta_{S\cup\{(q,r)\}} \triangleq F_{\hat{h}_{S\cup\{(q,r)\}}}(S\cup\{(q,r)\}) - F_{\hat{h}_S}(S\cup\{(q,r)\})$$

To guarantee that $F(S)$ is monotone non-decreasing, we require $\Delta_{S\cup\{(q,r)\}} \geq 0$ when adding a query-response pairs into S, there maybe no solution to such $F_{\hat{h}_{S\cup\{(q,r)\}}}(\cdot)$, and under this circumstance we remain $F(S)$ unchanged (i.e, set $F_{\hat{h}_{S\cup\{(q,r)\}}}(\cdot) = F_{\hat{h}_S}(\cdot)$).

In the other aspect, to make sure that $F(S)$ is submodular, the constraint is a little more complicate. we state it below and prove it in Lemma A.1.

Condition 4.6. (Condition for Submodularity).

$$\Delta_{S_t\cup\{(q,r)\}} \leq \delta_{S_{t-1}}(F_{\hat{h}_{S_{t-1}}}, (q,r)) - \delta_{S_t}(F_{\hat{h}_{S_{t-1}}}, (q,r))$$

, where $\delta_S(F,x)$ denotes $F(S\cup\{x\}) - F(S)$.

Algorithm 1: Worst Case Greedy Algorithm for EISSC

Input: H, Q, R, F, α
Output: S_T
begin
  $t \longleftarrow 0$;
  $S_0 \longleftarrow \emptyset$;
  $\hat{h}_{S_0} \longleftarrow \sum_{h_i\in H} \frac{1}{|H|} h_i$;
  while $F_{\hat{h}_{S_t}}(S_t) < \alpha$ **do**
    $\hat{q} \longleftarrow \arg\max_{q\in Q} \min_{r\in R}(F_{\hat{h}_{S_t}}(S_t\cup\{(q,r)\}) - F_{\hat{h}_{S_t}}(S_t))/c(q)$;
    Ask $\hat{q}$, receive $\hat{r}$;
    $S_{t+1} \longleftarrow S_t \cup \{(\hat{q},\hat{r})\}$;

$$\hat{h}_{S_{t+1}} \longleftarrow \begin{cases} \hat{h} = \underset{\hat{h}\in H^+}{\arg\min}\, G_{\hat{h}}(S_{t+1}) & \\ \quad \text{s. t.} 0 \leq F_{\hat{h}}(S_{t+1}) - F_{\hat{h}_{S_t}}(S_{t+1}) \leq \delta_t^{\triangle} & \text{if exists such } \hat{h}; \\ \hat{h}_{S_t} & \text{if no such } \hat{h} \end{cases}$$

    `/* if this process return a set` $\hat{H}$ `of candidates, set` $\hat{h} = \underset{h\in\hat{H}}{\arg\min}\, F_h(S_{t+1})$`, if there are still more than one condidates, select one randomly. */`
    $t \longleftarrow t+1$;
  end
end

Let $\delta_t^{\triangle}$ denotes $\delta_{S_{t-1}}(F_{\hat{h}_{S_{t-1}}}, (q,r)) - \delta_{S_t}(F_{\hat{h}_{S_{t-1}}}, (q,r))$, also, if $F(S)$ remains unchanged, due to the submodularity of $F_{\hat{h}_{S_{t-1}}}(\cdot)$, $\delta_t^{\triangle} \geq 0$ and $\Delta_{S_i\cup\{(q,r)\}} = 0$.

Using Definition 4.5, and constraints based on it we can then give the worst case greedy algorithm for ensemble interactive submodular set cover in Algorithm 1.

To guarantee the algorithm terminate within a bounded cost we require another condition which is reasonable in practice.

Condition 4.7. *There exists at least one non-zero term in sequence* $\langle \Delta_{S_i} \rangle_{i=1}^{T}$.

Theorem 4.8. *Given Condition 4.7, Algorithem 1 using Definition solves the finite hypothesis version of Problem 4.1 using cost at most* $GCC\Big((T-1)(\ln(\alpha(T-1)/\Delta)) - 1\Big) + \Delta_{S_T}$, *where* $\Delta = \sum_{i=1}^{T-1} \Delta_{S_i}$.

We prove this upper bound in Theorem A.3 and also breifly investigate the reduction of our ensemble method to classification problems in Appendix A.2 together with some discussion on computaional issues.

4.3 Infinite Ensemble Hypotheses Problem Tackling

Algorithm 2: The CG-Regression algorithm

Input: $S_i = \{(\boldsymbol{q}, r)\}, C, \nu \in (0, 1)$
Output: Linear combination from H
begin
  $t \longleftarrow 0$;
  while $\sum_i h_t(\boldsymbol{q}_i)(\lambda_i - \lambda_i^*) \geq 1$ **do**
    $t \longleftarrow t + 1$;
    Let $[\boldsymbol{\lambda}, \boldsymbol{\lambda}^*]$ be the solution of problem 2 using $t - 1$ hypotheses;
    $h_t \triangleq L(S_t, \mathbf{p})$, where $\mathbf{p} \triangleq \boldsymbol{\lambda} - \boldsymbol{\lambda}^*$;
  end
  Let $[\boldsymbol{\beta}, b]$ be the dual solution to $[\boldsymbol{\lambda}, \boldsymbol{\lambda}^*]$, i.e., a solution to problem 1;
  return $f = b + \sum_{i=1}^{t-1}(\beta_i h_i - \beta_i F_{h_i}(S_t))$;
end

For the classification case, there are always finite hypotheses in the augmented hypothesis class H^+ if the query class Q and the intended classes are finite. On the contrary, for regression the base hypothesis in H can even be infinite, and this causes a chanllenge that doing ensemble regression in infinite hypothesis space.

The method which our approach based on called semi-infinite linear programming(c.f., [8]), which means we consider a particular linear case of the EISSC described above. As the problem remains intractable if the specific form of F_H and G_h are not determined. Meanwhile, the discussion will be overly exhausted when nonlinear version of these two functions combined with semi-infinite programming.

We require linearity of F_h w.r.t the ensemble parameter β in hypothesis updating, which can be summarized in Assuming 4.9.

Assuming 4.9. (linearity of F_h w.r.t $\boldsymbol{\beta}$). $F_{\hat{h}} = \sum_{i=1}^{|H|} \beta_i F_{h_i}$.

Also, in regression we use a ε-insensitive loss (c.f., [8]) $l_\varepsilon(y, f(\boldsymbol{x})) = \max(0, y - f(\boldsymbol{x}) - \varepsilon)$ as G_h. Let S_t denote the set of current selected query-response pairs, and for each step of Algorithm 1, $F_{\hat{h}_{S_t}}(S_{t+1})$ and $\delta_t^{\triangle}$ is known constant. At last, in regression scenarios we replace quey-response pairs by input-output pairs, i.e., $\boldsymbol{q}_i$ as a input vector and r_i a response output.

With a l_1-norm, and applying a little transformation, the original ensemble procedure can be reformulated as a linear programming problem ((7) in [17] with two more constraints).

$$\begin{aligned} &\underset{\boldsymbol{\beta} \succeq 0, b, \boldsymbol{\xi} \succeq 0, \boldsymbol{\xi}^* \succeq 0}{\arg\min} \frac{C}{N}\left(\sum_{i=1}^{|S|+1} \xi_i + \xi_i^*\right) + \|\boldsymbol{\beta}\|_1 + C\nu\varepsilon \\ &\text{s. t.} \quad r_i - \hat{h}(\boldsymbol{q}_i) \leq \varepsilon + \xi_i, \ \ \hat{h}(\boldsymbol{q}_i) - r_i \leq \varepsilon + \xi_i^* \quad i = 1, \ldots, |H| \\ &\qquad\quad 0 \leq \sum_{i=1}^{|H|} \beta_i F_{h_i} - F_{\hat{h}_{S_t}}(S_{t+1}), \ \ \sum_{i=1}^{|H|} \beta_i F_{h_i} - F_{\hat{h}_{S_t}}(S_{t+1}) < \delta_t^{\triangle} \end{aligned} \tag{1}$$

where $\hat{h}(\boldsymbol{q}_i) \triangleq \sum_{i=1}^{|H|} \beta_i h_i(\boldsymbol{q}_i)$, $\nu \in (0, 1]$ is to help control the tube size ε.

The use of l_1-norm as regularizer aims to obtain sparse optimal solution. We show that the optimal number of hypotheses in the ensemble is not greater than the number of samples and independent of the size of a (finite) hypothesis space H (c.f., Corollary B.4).

Importing Lagrangian multipliers λ_i and λ_i^* to evaluate the error that targets go outside of the tube and construct the dual of LP-problem 1, we get

$$
\begin{aligned}
\underset{\boldsymbol{\lambda}\succeq 0,\boldsymbol{\lambda}^*\succeq 0}{\arg\min} & \sum_{i=1}^{|S|}(r_i + F_{\hat{h}_{S_t}}(S_{t+1}))(\lambda_i - \lambda_i^*) - \delta_t^{\triangle}\lambda_i^* \\
\text{s. t.} \quad & \sum_{i=1}^{|S_t|}\lambda_i - \lambda_i^* = 0, \ \ \sum_{i=1}^{|S_t|}\lambda_i + \lambda_i^* \le C\nu \\
& \lambda_i, \lambda_i^* \le C/N \quad i = 1, \dots, |S_i| \\
& \sum_{i=1}^{|S_t|}(h_k(\boldsymbol{q}_i) + F_{h_k}(S_t))(\lambda_i - \lambda_i^*) \le 1 \quad k = 1, \dots, |H|
\end{aligned}
\tag{2}
$$

Let h_k' denotes $h_k(\boldsymbol{q}_i) + F_{h_k}(S_t)$ and $p_i = \lambda_i - \lambda_i^*$, We can see the form of this problem is equivalent to the one discussed in [17]. We consider the magnitude of $\sum_i h(\boldsymbol{q}_i)pi$ as the measure of possiblity that the hypothesis can improve the ensemble. To extend the dual problem 2 to a infinite hypothesis set $\mathcal{H}$, As the solutionto the primal linear programming problem of any finite subset H of $\mathcal{H}$ is always primal feasible for arbitrary superset $H' \in \mathcal{H}$ of it. so the solutions that are simultaneously dual feasible to H' is also the optimal solution for the ensemble regression problem of H'. Define the base learning algorithm $L(S, \boldsymbol{p}) \triangle \underset{h\in\mathcal{H}}{\arg\max} \sum_{i=1}^{N} h(\boldsymbol{q}_i)p_n$, and we can immediately extend problem 2 to infinite hypothesis class version.

$$
\begin{aligned}
\underset{\boldsymbol{\lambda}\succeq 0,\boldsymbol{\lambda}^*\succeq 0}{\arg\min} & \sum_{i=1}^{|S|}(r_i + F_{\hat{h}_{S_t}}(S_{t+1}))(\lambda_i - \lambda_i^*) - \delta_t^{\triangle}\lambda_i^* \\
\text{s. t.} \quad & \sum_{i=1}^{|S_t|}\lambda_i - \lambda_i^* = 0, \ \ \sum_{i=1}^{|S_t|}\lambda_i + \lambda_i^* \le C\nu \\
& \lambda_i, \lambda_i^* \le C/N \quad i = 1, \dots, |S_i| \\
& \sum_{i=1}^{|S_t|} h_k'(\boldsymbol{q}_i)(\lambda_i - \lambda_i^*) \le 1 \quad \forall \boldsymbol{p} \in P
\end{aligned}
\tag{3}
$$

where P denotes the set of all dual feasible value of $\boldsymbol{p}$, and is equivalent to the compact polyhedron:

$$
P = \Big\{\boldsymbol{p} | \ \sum_{i=1}^{|S_t|}|p_i| \le C\nu, \sum_{i=1}^{|S_t|} p_i = 0, |p_i| \le C/N \text{ for } i = 1, \dots, |S_t|\Big\}
$$

Hence, we can conclude the requirements for H to make optimal solutions of 3 be finite.

Theorem 4.10. (Condition for Finite Optimal Solutions of Problem 3). *At t-th step of hypothesis updating, define $H' \triangleq \{h_i' \mid h_i' = h_i + F_{h_i}(S_t), h \in H\}$. If H' has the form of $\{h_i' \mid h_i' = L(S_t, \boldsymbol{p}_i), \boldsymbol{p}_i \in P\}$, there exists finite optimal solutions to the dual semi-infinite linear programming problem.*

We prove the Theorem 4.10 in Lemma B.7 and state the generic dual and pirmal semi-infinite linear programming in Appendix B.

Use the result above, we introduce a *Column Generation Algorithm* 2 to generate the best ensemble hypothesis each step, and then substitute it into Algorithm 1.It has been proved that it converges for SILP (c.f., Theorem 7.2 in [8]).

5 Simulation Experiments

Performance of Classification Version. With proper choice of F_h and G_h we can smoothly reduce our algorithm to a classification one as we discussed in Appendix A.2. We compared the classification version of our method with some baselines metioned in [7] and results are shown in figure 1. As we see, our approach has a remarkable improvement in achieving true utility with its cost remains still acceptable in a poor priori knowlegde scenario. This might because algorithms in previous work only concern more about learning than covering and is more likely to choose queries that can quickly increase G_h.

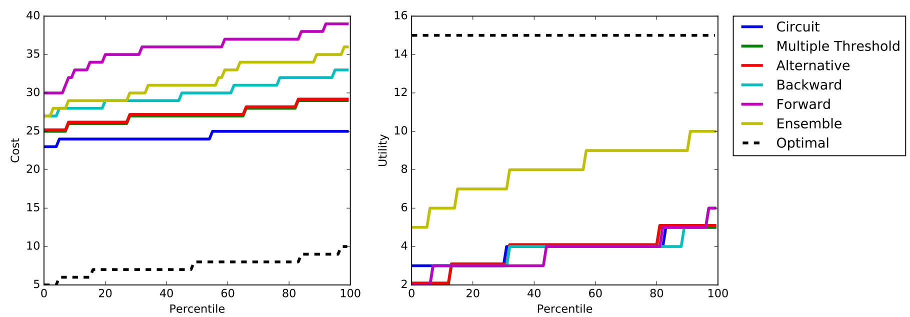

Figure 1: Comparison of classification version of EISSC against baselines in multiple threshold setting.

Validating for Regression Version We apply our method to both toy data and real data[2], and compare the performance under either $h^* \in H^+$ or $h^* \notin H^+$. As figure 2 shows, our method is promising for exactly obtaining utility under this interactive setting.

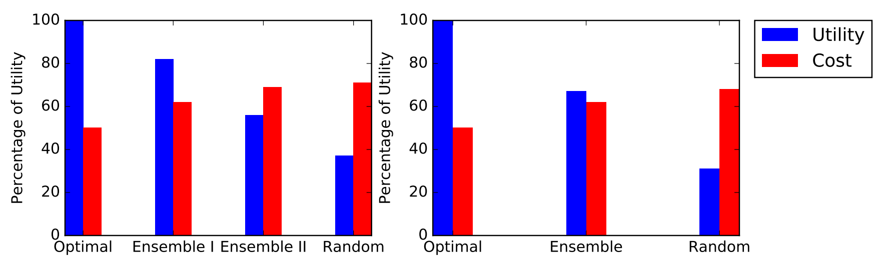

Figure 2: Comparison of Regression version of EISSC under two different settings.

6 Conclusions

We proposed an ensemble interactive submodular set cover algorithm which apply an ensemble method to all hypotheses we have every time we ask a question and receive a response. By greedy judgement based on this relatively more reliable hypothesis we obtain better performance in covering. We also discuss the infinite hypotheses problem and address a linear case of it.

References

[1] Ashwinkumar Badanidiyuru, Baharan Mirzasoleiman, Amin Karbasi, and Andreas Krause. Streaming submodular maximization: Massive data summarization on the fly. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 671–680. ACM, 2014.

[2] Yuri Y Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 105–112. IEEE, 2001.

[2] the Parkinsons Telemonitoring dataset [15] with manual modification, for concrete experiment settings please refer to Appendix C.2

[3] Satoru Fujishige. *Submodular functions and optimization*, volume 58. Elsevier, 2005.

[4] Andrew Guillory and Jeff A Bilmes. Interactive submodular set cover. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 415–422, 2010.

[5] Andrew Guillory and Jeff A Bilmes. Simultaneous learning and covering with adversarial noise. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 369–376, 2011.

[6] Andrew Russell Guillory. *Active Learning and Submodular Functions*. PhD thesis, 2012.

[7] Bryan D He and Yisong Yue. Smooth interactive submodular set cover. In *Advances in Neural Information Processing Systems*, pages 118–126, 2015.

[8] Rainer Hettich and Kenneth O Kortanek. Semi-infinite programming: theory, methods, and applications. *SIAM review*, 35(3):380–429, 1993.

[9] Stefanie Jegelka and Jeff Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1897–1904. IEEE, 2011.

[10] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[11] Pushmeet Kohli, M Pawan Kumar, and Philip HS Torr. P^3 & beyond: Move making algorithms for solving higher order functions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(9):1645–1656, 2009.

[12] Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability: Practical Approaches to Hard Problems*, 3:19, 2012.

[13] Andreas Krause and Carlos Guestrin. Near-optimal observation selection using submodular functions. In *AAAI*, volume 7, pages 1650–1654, 2007.

[14] Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 510–520. Association for Computational Linguistics, 2011.

[15] Max A Little, Patrick E McSharry, Eric J Hunter, Jennifer Spielman, and Lorraine O Ramig. Suitability of dysphonia measurements for telemonitoring of parkinson's disease. *Biomedical Engineering, IEEE Transactions on*, 56(4):1015–1022, 2009.

[16] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14(1):265–294, 1978.

[17] Gunnar Rätsch, Ayhan Demiriz, and Kristin P Bennett. Sparse regression ensembles in infinite and finite hypothesis spaces. *Machine Learning*, 48(1-3):189–218, 2002.

[18] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.

[19] Matthew Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Information Processing Systems*, pages 1577–1584, 2009.

[20] Laurence A Wolsey. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2(4):385–393, 1982.