Active Learning and Covering Problems with Precedence

author names withheld

Editor: Under Review for COLT 2026

Abstract

In the Bayesian Active Learning a hidden hypothesis is required to be uncovered. To do so, the learner is allowed to perform tests, each of which reveals partial information about the hidden hypothesis. Upon receiving this information, the learner adaptively selects the next test to be performed. The goal is to uncover the hidden hypothesis while performing as few tests as possible in the worst or average case.

In the covering problems, we are given a set of items and a collection of subsets that cover these items. The objective is to select a sequence of subsets that covers all items, which minimizing the worst or average covering cost.

For both types of problems, a natural constraint may arise that some tests can only be performed only after certain other tests (or some subsets can only be selected after selecting certain other subsets). We model such constraints using directed acyclic graphs (DAGs) that impose precedence on the tests or subsets. This paper explores the connection of active learning and covering problems under such constraints.

We show that given any bicriteria $(O(1)\,,\alpha)$-approximation ratio for the Precedence Constrained Set Cover, we can obtain an $O(\alpha \cdot \log n)$-approximation ratio for the Worst Case Active Learning with precedence constraints, where n is the number of hypothesis. Similarly, we prove that given any $O(\beta)$-approximation ratio for the Precedence Constrained Min-Sum Set Cover, we can obtain an $O(\beta \cdot \log n)$-approximation ratio for the Average Case Active Learning with Precedence Constraints. Finally, we provide several approximation algorithms for the Set Cover and Min-Sum Set Cover problems with various types of precedence constraints.

Keywords: Bayesian active learning, Set cover, Precedence constraints, Approximation Algorithms, Decision Trees

1. Introduction

Consider following problems:

- The *Precedence Constrained Bayesian Active Learning Problem* consists a set of $\mathcal{H}$ of n hypothesis, a set $\mathcal{T}$ of m tests and a DAG (directed acyclic graph) $\mathcal{F} = \{\mathcal{T}, \preceq\}$ encoding the precedence constraints between available tests. Among $\mathcal{H}$ a hidden hypothesis is required to be encovered. To do so, the learner is allowed to perform tests, each of which reveals partial information about the hidden hypothesis. Upon receiving this information, the learner adaptively selects the next test to be performed. Importantly, in order to perform such test the learner needs to perform all of its predecesors in $\mathcal{F}$ first. The goal is to uncover the hidden hypothesis while performing as few tests as possible. Depending on the chosen criterion we distinguish between the *Precedence Constrained Worst Case Active Learning* (PCWCAL) and *Precedence Constrained Average Case Active Learning* (PCACAL) problems.
- The *Precedence Constrained Covering Problem* consists of a set of n items $\mathcal{U}$, a collection $\mathcal{S}$ of m subsets of $\mathcal{U}$ that cover these items, and a DAG $\mathcal{F} = \{\mathcal{S}, \preceq\}$ encoding the precedence constraints between available subsets. The goal is to select a sequence of tests that covers at least K items. Depending on the chosen criterion we distinguish between the *Precedence Constrained Set Cover* (PCSC) and *Precedence Constrained Min-Sum Set Cover* (PCMSSC) problems. In the first we are only interested in minimizing the number of selected subsets, while in the second we want to minimize the average time it takes to cover an item.

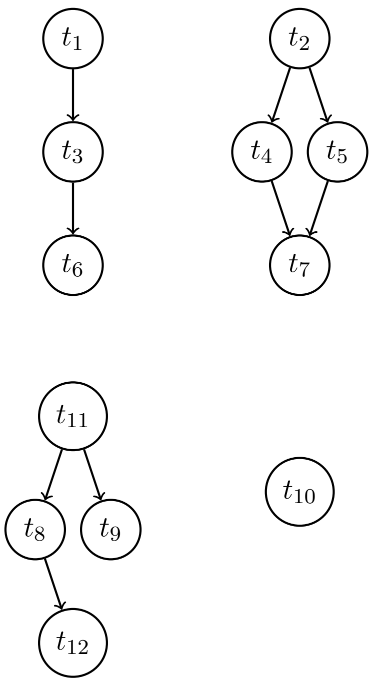
1.1. Our results and techniques

precedence/problem	**PCSC**	**PCMSSC**	**PCWCAL**	**PCACAL**
none	$O(\log n)$	4	$O(\log n)$	$O(\log n)$
inforest	$O(\log n)$*	4	$O(\log n)$*	$O(\log n)$*
outforest	$O(\log^2 n)$**	$O(\log n)$**	$O(\log^2 n)$*	$O(\log^2 n)$*
general	$O(\sqrt{n} \log n)$*	$O(\sqrt{n})$	$O(\sqrt{n} \log n)$*	$O(\sqrt{n} \log n)$*

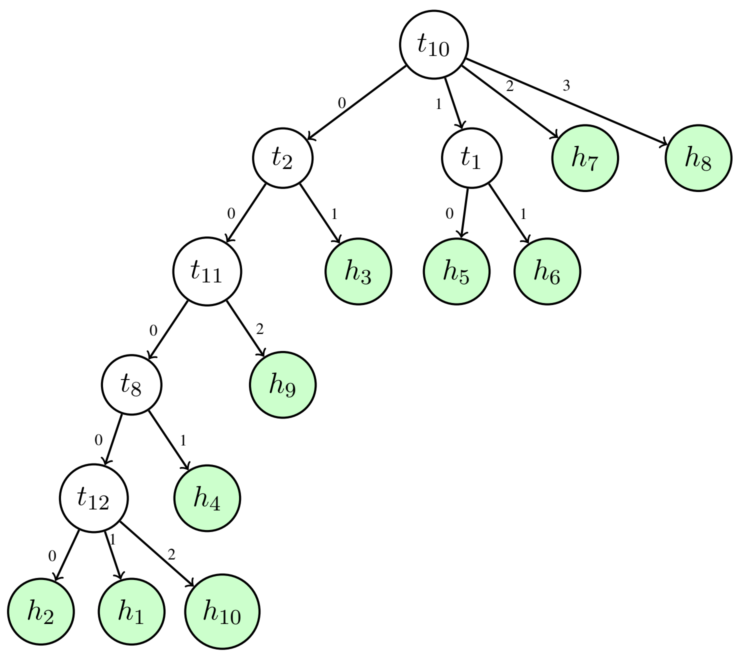
Table 1: Approximation algorithms for various covering and active learning problems under different precedence constraints. (* denotes new results, ** denotes previously unmentioned corollaries of known results)

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}
h_1	0	0	0	0	0	0	2	0	0	0	0	1
h_2	0	0	0	0	0	1	0	0	0	0	0	0
h_3	0	1	0	1	0	0	0	0	0	0	0	0
h_4	0	0	0	0	1	0	0	1	0	0	0	0
h_5	0	0	0	0	0	0	1	0	0	1	1	0
h_6	1	0	0	2	0	0	0	0	0	1	0	0
h_7	0	0	0	0	0	2	0	0	1	2	0	0
h_8	0	0	1	0	0	0	0	0	0	3	0	0
h_9	0	0	0	0	0	0	0	1	1	0	2	0
h_{10}	0	0	0	1	0	0	0	0	0	0	0	2

(a) Hypotheses and tests table

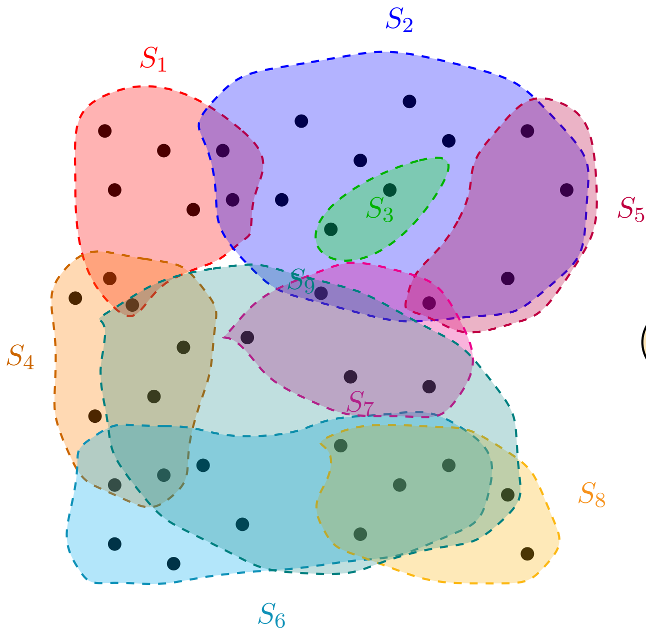


(b) Precedence DAG

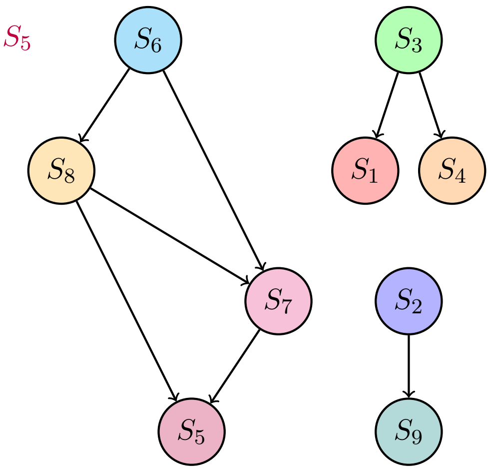


(c) Valid decision tree

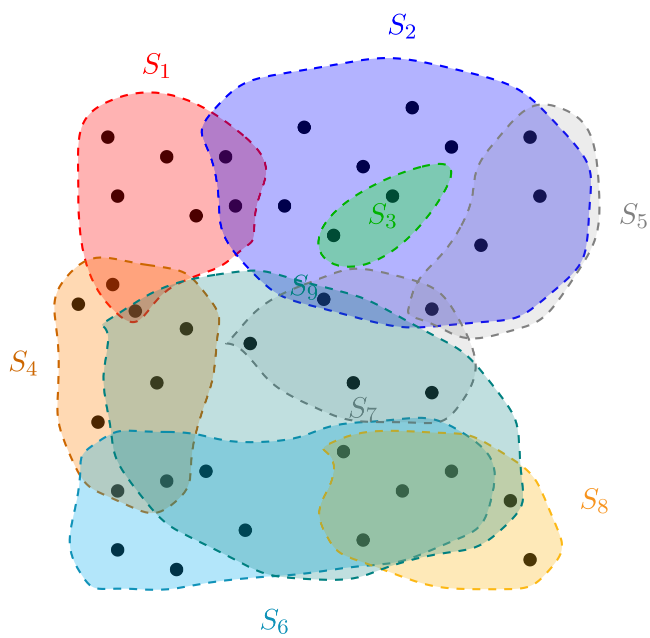
Figure 1: Example of a PCAL instance with 10 hypotheses and 12 tests. (a) Hypotheses-tests table. (b) Precedence DAG with four components. (c) A valid decision tree solution respecting precedence constraints.



(a) Universe and covering sets



(b) Precedence DAG



(c) Solution: selected sets $\{S_1, S_2, S_6, S_8, S_9\}$ (in color)

Figure 2: Example of a PCCP instance with 39 elements and 9 covering sets. (a) Universe with covering sets. (b) Precedence DAG with three components. (c) Solution using 7 selected sets (colored).

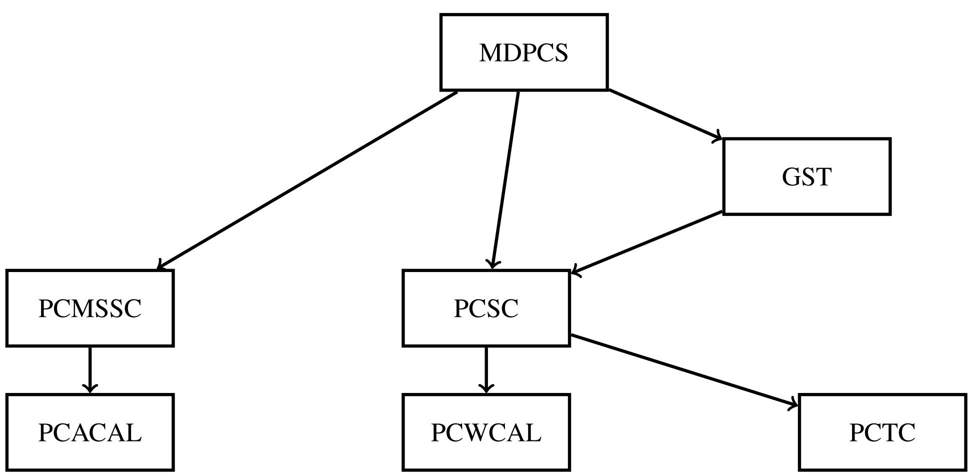


Figure 3: Relationships between covering and active learning problems, $\Pi_1 \to \Pi_2$ denotes that an approximation algorithm for problem Π_1 implies an approximation algorithm for problem Π_2.

2. Preliminaries

Definition 1 (Precedence constrained set cover (PCSC))

Definition 2 (Precedence constrained min-sum set cover (PCMSSC))

Definition 3 (Precedence constrained test cover (PCTC))

Definition 4 (Precedence constrained worst case active learning (PCWCAL))

Definition 5 (Precedence constrained average case active learning (PCACAL))

Definition 6 (Group Steiner Tree (GST))

Definition 7 (Max-Density Precedence-Closed Subfamily (MDPCS))

3. Active Learning via Covering Problems

We begin with the following folklore lemma concerning both worst and average case learning.

Lemma 8 *Let $I = (\mathcal{H}, \mathcal{T}, \mathcal{F})$ be any PCAL instance. Let $\mathcal{H}' \subseteq \mathcal{H}$. Then $\texttt{OPT}(\mathcal{H}', \mathcal{T}, \mathcal{F}) \leq \texttt{OPT}(I)$.*

3.1. Worst Case

Definition 9 (Pairsep) *Let D be any decision tree for the PCWCAL problem instance $\mathcal{I} = (\mathcal{H}, \mathcal{T}, \mathcal{F})$. We define a sequence of tests P_D called* pairsep *as follows. Initially, P_D is empty and $\mathcal{H}' = \mathcal{H}$. While $\binom{|\mathcal{H}'|}{2} > \binom{|\mathcal{H}|}{2}/2$, we append to P_D the test $r(D_{\mathcal{H}'})$ and update $\mathcal{H}'$ to be the set of hypotheses corresponding to the child of $D_{\mathcal{H}'}$ that contains the most hypotheses. If $\texttt{COST}(D) = \texttt{OPT}(\mathcal{I})$, then we denote $P^*(\mathcal{I}) = P_D$ (ties broken arbitrarily).*

It should be remarked that P_D is well-defined, as each test in P_D can have at most one child associated with more than half of the pairs hypotheses in $\mathcal{H}'$. Since P_D is a subpath of D, we also have the following simple observation.

Observation 1 *Let I be any instance of PCWCAL. Then $|P^*(I)| \leq \texttt{OPT}(I)$.*

This allows to use $|P^*(I)|$ as a lower bound on $\texttt{OPT}(I)$ in the analysis of the approximation algorithm for PCWCAL. We have the following lemma:

Lemma 10 *Let $I = (\mathcal{H}, \mathcal{T}, \mathcal{F})$ be any PCWCAL instance. Let S^* be the optimal solution for the PCSC on instance $(\mathcal{U}, \mathcal{T}, \mathcal{F})$ with $K = n/2$, where $\mathcal{U} = \{(h, j) \mid h, j \in \mathcal{H}\}$ and a test t covers (h, j) if it distinguishes h and j. Then, $|S^*| \leq |P^*(I)|$.*

Proof We show that $P^*(I)$ is a feasible solution for the PCSC instance $(\mathcal{U}, \mathcal{T}, \mathcal{F})$ with $K = n/2$. By definition, for every $\mathcal{H}' \in \mathcal{H} - P^*(I)$, we have $\binom{|\mathcal{H}'|}{2} \leq \binom{|\mathcal{H}|}{2}/2$. Therefore the number of pairs covered by tests in $P^*(I)$ is at least:

$$\binom{|\mathcal{H}|}{2} - \sum_{\mathcal{H}' \in \mathcal{H} - P^*(I)} \binom{|\mathcal{H}'|}{2} \geq \binom{|\mathcal{H}|}{2} - \frac{\binom{|\mathcal{H}|}{2}}{2} = \frac{\binom{|\mathcal{H}|}{2}}{2}.$$

Therefore, by the optimality of S^*, we have $|S^*| \leq |P^*(I)|$ as required. ■

Theorem 11 *If there is an (γ, α)-bicriteria approximation algorithm for PCSC then there is an $O\left(\frac{\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \log n\right)$-approximation algorithm for PCWCAL. In particular when $\gamma = O(1)$, the approximation is $O(\alpha \cdot \log n)$.*

Proof The following observation follows by Lemmas 8 and 10:

Observation 2 *Let D_S be the decision tree built on tests from S closed under $\mathcal{F}$. Then, $\texttt{COST}(D_S) \leq \alpha \cdot |P^*(I)|$.*

Algorithm 1: The $O(\alpha \cdot \log n)$-approximation algorithm for the PCWCAL

procedure WORSTDECISIONTREE($\mathcal{H}, \mathcal{T}, \mathcal{F}$)

```
    U ← {(h, j) | h, j ∈ H}
    foreach t ∈ T do
        Mark t as covering (h, j) ∈ U if t distinguishes h and j
    end
    S ← Run the α-approximation algorithm for PCSC on instance (U, T, F) with K = n/2
    D ← D_S ← any decision tree built on tests from S closed under F
    foreach H' ∈ H − S do
        D' ← WORSTDECISIONTREE(H', T − S, F − S)
        Attach D' to the leaf of D corresponding to H'
    end
    return D
```

We are now ready to prove the theorem.

Lemma 12 *Let D be the decision tree returned by* WORSTDECISIONTREE *on input $I = (\mathcal{H}, \mathcal{T}, \mathcal{F})$. Then, $\mathit{COST}(D) \leq \frac{2\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \log n \cdot \mathit{OPT}(I)$.*

Proof We prove the lemma by induction on $p = \binom{|\mathcal{H}|}{2}$. The base case when $n = 1$ and $p = 2$ is trivial since there are no pairs to cover. Assume by induction that for every $I' = (\mathcal{H}', \mathcal{T}, \mathcal{F})$ such that $\mathcal{H}' \in \mathcal{H} - S$ and $n' = |\mathcal{H}'|$ we have $\mathtt{COST}(D') \leq \frac{\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \log \binom{n'}{2} \cdot \mathtt{OPT}(I')$, where D' is the decision tree returned by WORSTDECISIONTREE on input I'. We have that:

$$
\begin{aligned}
\mathtt{COST}(D) \leq& \mathtt{COST}(D_S) + \max_{\mathcal{H}' \in \mathcal{H} - S} \mathtt{COST}\big(D'\big) \\
\leq& \alpha \cdot |S^*| + \max_{\mathcal{H}' \in \mathcal{H} - S} \frac{\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \log \binom{n'}{2} \cdot \mathtt{OPT}\big(I'\big) \\
\leq& \alpha \cdot |P^*(I)| + \frac{\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \log\left(\left(\frac{2\gamma-1}{2\gamma}\right) \cdot \binom{n}{2}\right) \cdot \mathtt{OPT}(I) \\
=& \alpha \cdot \mathtt{OPT}(I) + \frac{\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \left(\log \binom{n}{2}\right) \cdot \mathtt{OPT}(I) - \alpha \cdot \mathtt{OPT}(I) \\
=& \frac{\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \log \binom{n}{2} \cdot \mathtt{OPT}(I)
\end{aligned}
$$

Since $\log \binom{n}{2} \leq 2 \log n$, the lemma follows. ■

■

3.2. Average Case

Theorem 13 *If there is a β- approximation algorithm for PCMSSC then there is an $O(\beta \cdot \log n)$-approximation algorithm for PCACAL.*

Algorithm 2: The $O(\beta \cdot \log n)$-approximation algorithm for the PCACAL

procedure AVERAGEDECISIONTREE$(\mathcal{H}, \mathcal{T}, \mathcal{F})$

```
    𝒰 ← ℋ
    foreach t ∈ 𝒯 do
        Set t to cover u ∈ 𝒰 if for u ∈ U_{t,j}, |U_{t,j}| ≤ (3/4) · |U|
    end
    S ← Run the α-approximation algorithm for PCMSSC on instance (𝒰, 𝒯, ℱ) with K = n/4
    D ← D_S ← any decision tree built on tests from S closed under ℱ
    foreach ℋ' ∈ ℋ − S do
        D' ← AVERAGEDECISIONTREE(ℋ', 𝒯 − S, ℱ − S)
        Attach D' to the leaf of D corresponding to ℋ'
    end
    return D
```

4. Set covering with constraints

5. Hardness

6. Conclusions and Future Work

Appendix A. My Proof of Theorem 1

This is a boring technical proof.

Appendix B. My Proof of Theorem 2

This is a complete version of a proof sketched in the main text.