Minimum entropy Submodular Set Cover

Gabriel Istrate,Cosmin Bonchis

25 November 2010

eAustriaResearchInstitute,
Bd.V.Parvan no.4,cam.045B,Timisoara,RO-300223,Romania,
email: gabrielistrate@acm.org,cbonchis@ieat.ro

Abstract

Building on the approach in [Halperin and Karp(2005)],[Cardinal et al.(2008)Cardinal, Fiorini, and Joraet], we introduce and analyze a minimum entropy version of the *submodular set cover problem* [Wolsey(1982)], [Fujita(2000)], [Bar-Ilan et al.(2001)Bar-Ilan, Kortsarz, and Peleg]. We give a polynomial time additive approximation algorithm and show that this constant is the best possible unless $P = NP$.

1 Definition

Definition 1 *A function $f : \mathcal{P}(U) \rightarrow \mathbf{Z}_{+}$ is*

1. monotone *if $f(S) \leq f(T)$ whenever $S \subseteq T \subseteq U$.*
2. submodular *if $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$ for all $S, T \subseteq U$.*

Definition 2 ***[SUBMODULAR SET COVER]:***

1. *[GIVEN:] A set U and a monotone, submodular function $f : \mathcal{P}(U) \rightarrow \mathbf{Z}_{+}$ and a* cost function *$c : U \rightarrow \mathbf{R}_{+}$. The cost function satisfies the normalization condition $\sum_x c(x) = 1$.*
2. *[TO FIND:] A subset $S \subseteq U$ with $f(S) = f(U)$ (such a set is called* feasible*) of minimum cost. The cost of a set S, denoted $c(S)$, is simply the sum of costs of its elements.*

The performance of the Greedy algorithm for submodular set cover was studied by Wolsey [Wolsey(1982)], who showed that results for set cover generalize to this setup. Generalizations were given, among other papers, in [Bar-Ilan et al.(2001)Bar-Ilan, Kortsarz, and P

The main interest of this paper is the approximability of the following problem:

Definition 3 ***[MINIMUM ENTROPY SUBMODULAR SET COVER]:***

1. *[GIVEN:] A set U and a monotone, submodular function $f : \mathcal{P}(U) \rightarrow \mathbf{Z}_+$.*

2. *[TO FIND:] A subset $S \subseteq U$ with $f(S) = f(U)$ (such a set is called* feasible*), and a distribution $D = (p_e)_{e \in S}$ that minimizes the entropy*

$$H(D) = -\sum_e p_e \log(p_e).$$

The distribution D is subject to the following two constraints:

   (a) $p_e = 0$ for all $e \notin S$.

   (b) For all $T \subseteq S$ $0 \leq \sum_{e \in T} p_e \leq \frac{f(T)}{f(U)}$.

Example 1 *Let $X = \{1, 2, \ldots, n\}$ for some $n \geq 1$ and $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ be a family of subsets of X which covers X. Associate to instance $(X, \mathcal{P})$ of the Minimum Entropy Set Cover problem [Halperin and Karp(2005)] the following set system:*

1. *$U = \{1, 2, \ldots, m\}$.*
2. *For $S \subseteq U$ define $X_S = \bigcup_{i \in S} P_i$ and $f(S) = |X_S|$.*
3. *Finally, $c(i) = 1$ for all $i \in U$.*

Example 2 *Further generalize the example above to the case of the Minimum Entropy Weighted Set Cover, defined informally in [Cardinal et al.(2008)Cardinal, Fiorini, and Joraet]. Speccifically, Let $X = \{1, 2, \ldots, n\}$ for some $n \geq 1$, let $(b_\nu)_{\nu \in X}$ be a system of positive weights on X and $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ be a family of subsets of X which covers X.*

Associate to instance $(X, \mathcal{P})$ of the Minimum Entropy Weighted Set Cover problem the following set system:

1. *$U = \{1, 2, \ldots, m\}$.*
2. *For $S \subseteq U$ define $f(S) = \sum_{x \in (\bigcup_{i \in S} P_i)} b_x$.*
3. *Finally, $c(i) = 1$ for all $i \in U$.*

2 The Greedy Algorithm for Minimum Entropy Submodular Set Cover

Algorithm 1 GREEDY MESSC:

INPUT: A triplet (U, f)
  $A := \emptyset$;
  While (there exists $e \in U$ with $\Delta_e f(A) > 0$)
    choose $e \in U$ to maximize $\Delta_a f(A)$;
    $p_e = \frac{\Delta_e f(A)}{f(U)}$;
    $A := A \cup \{e\}$;
OUTPUT: Distribution $D = (p_e)_{e \in A}$.

The intuition is that the probability distribution generated by the Greedy algoritm is minimizing the entropy of submodular set cover problem.

References

[Bar-Ilan et al.(2001)Bar-Ilan, Kortsarz, and Peleg] J. Bar-Ilan, G. Kortsarz, and D. Peleg. Generalized submodular cover problems and applications. *Theoretical Computer Science*, 250(1–2):179–200, 2001.

[Cardinal et al.(2008)Cardinal, Fiorini, and Joraet] J. Cardinal, S. Fiorini, and G. Joraet. Tight results on minimum entropy set cover. *Algorithmica*, 51 (1):49–60, 2008.

[Fujita(2000)] T. Fujita. Approximation algorithms for submodular set cover with applications. *IEICE Transactions on Information and Systems*, E83-D (3):480–487, 2000.

[Halperin and Karp(2005)] E. Halperin and R. Karp. The minimum entropy set cover problem. *Theoretical Computer Science*, 348(2–3):340–350, 2005.

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