J Glob Optim (2011) 50:229–234
DOI 10.1007/s10898-010-9563-3

On minimum submodular cover with submodular cost

Hongjie Du · Weili Wu · Wonjun Lee ·
Qinghai Liu · Zhao Zhang · Ding-Zhu Du

Received: 9 June 2010 / Accepted: 15 June 2010 / Published online: 26 June 2010

Abstract In this paper, we show that a minimum non-submodular cover problem can be reduced into a problem of minimum submodular cover with submodular cost. In addition, we present an application in wireless sensor networks.

Keywords Minimum submodular cover · Submodular cost · MOC-CDS

1 Introduction

Consider the following hitting set problem:

> HITTING SET: Given a collection $\mathcal{C}$ of subsets of a finite set E, find a minimum subset A of E such that for every $S \in \mathcal{C}$, $S \cap A \neq \emptyset$.

H. Du · W. Wu · D.-Z. Du
Department of Computer Science, University of Texas at Dallas, Richardson, TX 75080, USA
e-mail: hongjiedu@utdallas.edu

W. Wu
e-mail: weiliwu@utdallas.edu

D.-Z. Du
e-mail: dzdu@utdallas.edu

W. Lee (✉)
Department of Computer Science and Engineering, Korea University, Seoul, Republic of Korea
e-mail: wlee@korea.ac.kr

Q. Liu · Z. Zhang
College of Mathematics and System Sciences, Xinjiang University, Urumqi, Xinjiang, 830046
People's Republic of China
e-mail: liuqh506@163.com

Z. Zhang
e-mail: zhzhao@xju.edu.cn

For every $A \subseteq E$, define

$$f(A) = |\{S \in \mathcal{C} \mid S \cap A \neq \emptyset\}|.$$

Then f has the following properties:

1. $f(\emptyset) = 0$.
2. f is *monotone increasing*, i.e., $A \subset B \Rightarrow f(A) \leq f(B)$.
3. f is *submodular*, i.e., for any two subsets A and B of E,

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B).$$

A function satisfying above three conditions is also called a *polymatroid function* on 2^E. Therefore, f is a polymatroid function on 2^E. With the function f, Hitting Set can be represented as a Minimum Modular Cover problem as follows:

$$\min\{|A| \mid f(A) = f(E), A \subseteq E\}.$$

Wolsey [1] showed that if f is an integer, then the Minimum Modular Cover problem has a greedy approximation with the performance ratio $H(\gamma)$ where $H(\gamma) = \sum_{i=1}^{\gamma} \frac{1}{i}$ is called the harmonic function, and $\gamma = \max\{f(\{x\}) \mid x \in E\}$.

Next, we consider a generalization of the Hitting Set:

Generalized Hitting Set: Given m nonempty collections $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m$ of subsets of a finite set E, find the minimum subset A of E such that every $\mathcal{C}_i$ has a member $S \subseteq A$.

Let us define the function f on 2^E in a similar way:

$$f(A) = |\{\mathcal{C}_i \mid \mathcal{C}_i \text{ has a member } S \subseteq A\}|.$$

Clearly, $f(\emptyset) = 0$ and f is monotone increasing. Therefore, we can also formulate the problem as follows:

$$\min\{|A| \mid f(A) = f(E), A \subseteq E\}.$$

However, in general, f is not submodular. To see this, let us show a counter example:

Let $E = \{a, b, c\}$, $\mathcal{C}_1 = \{\{a\}\}$, $\mathcal{C}_2 = \{\{b, c\}\}$, $A = \{a, b\}$ and $B = \{a, c\}$. Then

$$f(A) + f(B) = 2 < f(A \cup B) + f(A \cap B) = 3.$$

Since f is not submodular, we can not apply the theorem of Wolsey [1] to obtain a greedy approximation for the Generalized Hitting Set. What should we do? A possibility is to employ the methods in [2].

In this paper, we propose a new method by reducing it into a minimum submodular cover with submodular cost [3] in Sect. 2. We also present an application in Sect. 3. Finally, we conclude the paper in Sect. 4.

2 Minimum submodular cover with submodular cost

Let $\mathcal{C} = \cup_{i=1}^{m} \mathcal{C}_m$. For every subcollection $\mathcal{A} \subseteq \mathcal{C}$, define

$$f(\mathcal{A}) = |\{\mathcal{C}_i \mid \mathcal{A} \cap \mathcal{C}_i \neq \emptyset\}|$$

and

$$c(\mathcal{A}) = |\cup_{A \in \mathcal{A}} A|.$$

Lemma 1 *f and c are polymatroid functions on $2^{\mathcal{C}}$.*

Proof Clearly, $f(\emptyset) = c(\emptyset) = 0$, and both f and c are monotone increasing. Next, we show that they are submodular. For any two subcollections $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{C}$,

$$\begin{aligned} f(\mathcal{A}) + f(\mathcal{B}) &= f(\mathcal{A} \cup \mathcal{B}) + |\{\mathcal{C}_i \mid \mathcal{C}_i \cap \mathcal{A} \neq \emptyset \text{ and } \mathcal{C}_i \cap \mathcal{B} \neq \emptyset\}| \\ &\geq f(\mathcal{A} \cup \mathcal{B}) + |\{\mathcal{C}_i \mid \mathcal{C}_i \cap (\mathcal{A} \cap \mathcal{B}) \neq \emptyset\}| \\ &= f(\mathcal{A} \cup \mathcal{B}) + f(\mathcal{A} \cap \mathcal{B}). \\ c(\mathcal{A}) + c(\mathcal{B}) &= c(\mathcal{A} \cup \mathcal{B}) + |(\cup_{A \in \mathcal{A}} A) \cap (\cup_{B \in \mathcal{B}} B)| \\ &\geq c(\mathcal{A} \cup \mathcal{B}) + | \cup_{A \in \mathcal{A} \cap \mathcal{B}} A| \\ &= c(\mathcal{A} \cup \mathcal{B}) + c(\mathcal{A} \cap \mathcal{B}). \end{aligned}$$

□

Lemma 2 *The* GENERALIZED HITTING SET *is equivalent to the following problem:*

$$\min\{c(\mathcal{A}) \mid f(\mathcal{A}) = f(\mathcal{C}), \mathcal{A} \subseteq \mathcal{C}\}. \quad (1)$$

Proof $f(\mathcal{A}) = f(\mathcal{C})$ is equivalent to the condition that for every $i = 1, 2, \ldots, m$, $\mathcal{C}_i \cap \mathcal{A} \neq \emptyset$.

For the minimum solution A of the GENERALIZED HITTING SET, suppose A_i is a subset of A such that $A_i \in \mathcal{C}_i$. Then we must have $A = \cup_{i=1}^m A_i$, i.e., $|A| = c(\{A_1, A_2, \ldots, A_m\})$.

$$|A| \geq \min\{c(\mathcal{A}) \mid \mathcal{A} \subseteq \mathcal{C}\}.$$

Conversely, suppose $\mathcal{A}$ is a minimum solution of the problem (1). Set $A = \cup_{A \in \mathcal{A}} A$. Then $|A| = c(\mathcal{A})$. Thus, for any minimum solution A of the GENERALIZED HITTING SET,

$$|A| = \min\{c(\mathcal{A}) \mid \mathcal{A} \subseteq \mathcal{C}\}.$$

Moreover, from a minimum solution A of the GENERALIZED HITTING SET, we can easily compute a minimum solution $\{A_1, A_2, \ldots, A_m\}$ of problem (1), vice versa. Therefore, these two problems are equivalent. □

The problem (1) is a minimum submodular cover with submodular cost studied by Wan et al. [3]. From their results and by Lemmas 1 and 2, we can obtain the following theorem:

Theorem 3 *Let $\mathcal{C} = \cup_{i=1}^m \mathcal{C}_i$. Suppose f and c are polymatroid functions and f is an integer. Then Greedy Approximation GHS* (Algorithm 1) *for* GENERALIZED HITTING SET *has the performance ratio $\chi H(\gamma)$, where*

$$\gamma = \max\{f(\{S\}) \mid S \in \mathcal{C}\}$$

and

$$\chi = \min\left\{ \frac{\sum_{S \in \mathcal{A}} c(\{S\})}{c(\mathcal{A})} \,\middle|\, \mathcal{A} \text{ is a minimum solution} \right\}.$$

3 Application

When all sensors are identical, the wireless sensor network can be formulated as a unit disk graph $G = (V, E)$, in which all nodes lie in the Euclidean plane and an edge (u, v) exists if and only if the distance between u and v is at most one.

```
A ← ∅;
while ∃S ∈ C such that Δ_S f(A) > 0 do
    select S ∈ C to maximize Δ_S f(A) / c({S});
    A ← A ∪ {S};
Output A.
```

Algorithm 1: Greedy algorithm for GENERALIZED HITTING SET

A subset $D \subseteq V$ is called a *dominating set* if every node not in D is adjacent to a node in D. If a connected subgraph is induced by D, then D is called a *connected dominating set*. The connected dominating set plays an important role in wireless network where it has another name *virtual backbone* [4–6].

To reduce routing cost [7] and to balance road load [8], Ding et al. [7] and Willson et al. [8] studied the following problem:

> MOC-CDS: Given a graph, find the minimum connected dominating set D such that for any two nodes u and v, there exists a shortest path between them and with all intermediate nodes in D.

Ding et al. [7] showed that MOC-CDS is NP-hard and has no polynomial-time approximation algorithm with the performance ratio $\rho \ln \delta$ for $0 < \rho < 1$ unless $NP \subseteq DTIME(n^{\log \log n})$ where δ is the maximum node degree of the input graph. They also designed a polynomial-time distributed algorithm which produces approximation solution within a factor of $H(\frac{\delta(\delta-1)}{2})$ from optimal. Could MOC-CDS have a polynomial-time constant-approximation? This is an open problem. In the following, we approach to the solution of this open problem by showing a little weaker result. We will construct a connected dominating set D of size within a constant factor from $opt_{MOC-CDS}$, the optimal of MOC-CDS, and with a property that for any two nodes u and v, there is a path with all intermediate nodes in D and the length within a factor of three from the shortest path.

Let $d(u, v)$ denote the length of the shortest path between u and v and $d_D(u, v)$ the minimum length of path connecting u and v through D, i.e., with all intermediate nodes in D. We show a property of $d_D(u, v)$.

Lemma 4 *If for any two nodes u and v with a distance two, $d_D(u, v) \leq 6$, then for any two nodes u and v, $d_D(u, v) \leq 5 \cdot d(u, v)$.*

Proof Suppose $(u, w_1, w_2, \ldots, w_k, v)$ is a shortest path between u and v where $k = d(u, v) - 1$. Since $d(u, w_2) = d(w_2, w_4) = \cdots = 2$, there exists paths $(u, x_{11}, \ldots, x_{15}, w_2)$, $(w_2, x_{31}, \ldots, x_{35}, w_4), \ldots$ such that all intermediate nodes $x_{11}, \ldots, x_{15}, x_{31}, \ldots, x_{35}, \ldots$ belong to D. Moreover, since $d(x_{15}, x_{31}) = d(x_{35}, x_{51}) = \cdots = 2$, there exists paths $(x_{15}, x_{21}, \ldots, x_{25}, x_{31})$, $(x_{35}, x_{41}, \ldots, x_{45}, x_{51}), \ldots$ such that all intermediate nodes $x_{21}, \ldots, x_{25}, x_{41}, \ldots, x_{45}, \ldots$ belong to D. Therefore, we obtain a path between u and v with $5k$ intermediate nodes x_{ij}, $1 \leq i \leq k$ and $1 \leq j \leq 5$, all in D. Therefore,

$$d_D(u, v) \leq 5k + 1 = 5(d(u, v) - 1) + 1 < 5d(u, v).$$

□

Next, we construct a connected dominating set for the unit disk graph $G = (V, E)$. At the first stage, we construct a maximal independent set I. From a result of Li et al. [9], we have

$$|I| \leq \frac{11}{3} opt_{MOC-CDS} + \frac{4}{3}. \tag{2}$$

At the second stage, we connect I into a connected dominating set D such that

(a) $|D| \leq \alpha \cdot opt_{MOC-CDS}$ for a constant α, and
(b) $d_D(u, v) \leq 5d(u, v)$ for $u, v \in V$.

To do so, we reduce this duty to the Generalized Hitting Set problem as follows:

For any two nodes u and v in I with a distance at most four, denoted as $\mathcal{C}_{uv}$, the collection of all subsets consists that all intermediate nodes on a path connecting u and v with at most three intermediate nodes.

Lemma 5 *If $A \subseteq V$ hits every $\mathcal{C}_{uv}$ for every two nodes u, v in I with a distance at most four, then*

$$d_D(x, y) \leq 5d(x, y)$$

for every two nodes x, y in V where $D = I \cup A$.

Proof For any node x, let x' denote a node in I, which is either x or a node adjacent to x. Consider any two nodes $x, y \in V$ with a distance two. Then $d(x', y') \leq 4$. Therefore, $d_A(x', y') \leq 4$. Hence, $d_D(x, y) \leq 6$. By Lemma 4, for every two nodes x, y in V, $d_D(x, y) \leq 5d(x, y)$. □

Now, we employ the Greedy Algorithm GHS to find an approximation solution A. Its size is estimated as follows:

Lemma 6 *Let A be obtained by Greedy Algorithm GHS for hitting all $\mathcal{C}_{uv}$ over all pairs of nodes u, v in I with a distance at most four. Then*

$$|A| \leq 545 \cdot H(25) \cdot opt_{MOC-CDS}.$$

Proof First, we note that Wan et al. [6] showed that each node is adjacent to at most five nodes in I. Therefore, there are at most 25 paths sharing the same intermediate nodes and with endpoints in I. This means $\gamma = \max_{S \in \mathcal{C}} f(\{S\}) \leq 25$.

To estimate χ, we first estimate the number of nodes in I within a distance d from a node u. For each of such nodes, we draw a disk with center at the node and radius 0.5. Then all such disks are disjoint, contained in a disk with center at u and radius $d + 0.5$. Therefore, the number of such nodes in I is at most $\frac{\pi(d+0.5)^2}{\pi 0.5^2} = (2d + 1)^2$.

Now, suppose u is in the intermediate of a path connecting x and y in I with at most three intermediate nodes. Then there are two cases.

Case 1 Both x and y are within a distance at most two from u. The number of pairs $\{x, y\}$ in this case is at most $25(25 - 1)/2 = 300$.

Case 2 One of x and y has a distance three from u. Then the other must have at most a distance one from u. Therefore, the number of pairs $\{x, y\}$ in this case is at most $5 \cdot 7^2 = 245$.

Putting two cases together, we obtain $\chi \leq 545$. Moreover, any feasible solution of MOC-CDS is a feasible solution of the Generalized Hitting Set with those collections $\mathcal{C}_{uv}$. Therefore, $|A| \leq 545 \cdot H(25) \cdot opt_{MOC-CDS}$. □

Now, we have our final conclusion.

Theorem 7 *Given any connected unit disk graph $G = (V, E)$, a connected dominating set D can be constructed in polynomial-time such that*

$$|D| \leq \left(545H(25) + \frac{11}{3}\right) opt_{MOC-CDS} + \frac{4}{3}$$

and

$$d_D(u, v) \leq 5d(u, v)$$

for any two nodes $u, v \in V$.

Proof If G is a complete graph, then any point forms a connected dominating set meeting our requirements. Otherwise, any maximal independent set I contains at least two nodes. Note that I has the property that for any partition (I_1, I_2) of I, there exist $x \in I_1$ and $y \in I_2$ such that $d(x, y) \leq 3$. Therefore, we would construct a path connecting x and y at the second stage. Let A be obtained at the second stage. Then $D = I \cup A$ meets our requirement. □

4 Conclusion

In this paper, we show that a minimum non-submodular cover problem can be reduced into a problem of minimum submodular cover with submodular cost. We proposed a greedy algorithm for Generalized Hitting Set with the performance ratio $\chi H(\gamma)$, where $\gamma = \max\{f(\{S\} \mid S \in \mathcal{C}\}$ and $\chi = \min\left\{\frac{\sum_{S \in \mathcal{A}} c(\{S\})}{c(\mathcal{A})} \mid \mathcal{A} \text{ is a minimum solution}\right\}$. In addition, we present an application in MOC-CDS problem in wireless sensor networks.

Acknowledgments This work was supported in part by National Science Foundation of USA under grants CCF-0728851 and CCF-9208913. This research was also jointly supported by MEST, Korea, under WCU (R33-2008-000-10044-0), a KOSEF Grant funded by the Korean Government (MEST) (No. R01-2007-000-11203-0), a KRF Grant (KRF-2008-314-D00354), and MKE, Korea under ITRC NIPA-2010-(C1090-1021-0008). Corresponding author: Wonjun Lee.

References

1. Wolsey, L.A.: An analysis of the greedy algorithm for submodular set covering problem. Combinatorica **2**(4), 385–393 (1982)
2. Du, D.-Z., Graham, R.L., Pardalos, P.M., Wan, P.-J., Wu, W., Zhao, W.: Analysis of greedy approximations with nonsubmodular potential functions. In: Proceedings of the 19th annual ACM-SIAM symposium on dicrete algorithms (SODA), pp. 167–175 (2008)
3. Wan, P.-J., Du, D.-Z., Pardalos, P.M., Wu, W.: Greedy approximations for minimum submodular cover with submodular cost. Comput. Optim. Appl. **45**(2), 463–474 (2010)
4. Cardei M., Cheng M.X., Cheng X., Du D.-Z.: Connected domination in ad hoc wireless networks. In: Proceedings of the 6th international conference on computer science and informatics (2002)
5. Sivakumar, R., Das, B., Bharghavan, V.: An improved spine-based infrastructure for routing in ad hoc networks. In: IEEE Symposium on Computer and Communications (1998)
6. Wan, P.-J., Alzoubi, K. M., Frieder, O.: Distributed construction of connected dominating set in wireless ad hoc networks. In: Proceedings of 3rd ACM international workshop on discrete algorithms and methods for mobile computing and communications, pp. 7–14 (1999)
7. Ding, L., Gao, X., Wu, W., Lee, W., Zhu, X., Du, D.-Z.: Distributed construction of connected dominating sets with minimum routing cost in wireless network. In: The 30th international conference on distributed computing systems (2010)
8. Willson, J., Gao, X., Qu, Z., Zhu, Y., Li, Y., Wu, W.: Efficient distributed algorithms for topology control problem with shortest path constraints, discrete mathematics. Algorithms Appl. **1**(4), 437–461 (2009)
9. Wan, P.-J., Wang, L., Yao, F.: Two-phased approximation algorithms for minimum CDS in wireless ad hoc networks. In: ICDCS, pp. 337–344 (2008)