A Weighted Set Cover Model for Task Planning of Earth Observation Satellites

Pengyun Li, Hongyan Li(✉), and Jun Chang

State Key Laboratory of Integrated Service Networks, Xidian University, Xian 710071, China
hyli@xidian.edu.cn

Abstract. Due to the diversification of observation missions and the differentiation of satellite resources, the task scheduling of Earth observation satellites has always been an NP-hard problem. In this paper, aiming at multi-load Earth observation satellite mission scheduling, considering multi-satellite coordinated observation, facing regional target mission and point target mission, a weighted set cover model is proposed to represent the coupling relationship between multi-satellite and multi-task. The classical greedy approximation algorithm is used to optimize the sum of satellite observation time windows. The model-based algorithm can effectively save satellite storage resources and sensor resources, and realize multi-satellite coordinated observation task scheduling.

Keywords: Task planning · Set cover model · Greedy approximation algorithm

1 Introduction

As an effective tool for the exploring of the earth's resources, the earth observation satellites (EOS) have been widely used in the fields of the agricultural monitoring, the natural disaster warning, the large-scale infrastructure construction and the ground military target identification [12]. However, with the increase of the number of observing tasks and the diversification of observational demands, the scarce resources on the star become more and more valuable. How to maximize the resource utilization through reasonable scheduling has become an important issue to be solved urgently.

For example, when it is necessary to observe a large range of targets, or to make continuous observations on a hot-spot area, it will take a long time or cannot complete at all depending on the periodic motion of one single satellite, where the joint observation of multiple satellites and the splitting-aggregation of multiple tasks will greatly shorten the observation delay and save the satellite resources.

Supported by organization x.

M. Jia et al. (Eds.): WiSATS 2019, LNICST 280, pp. 449–455, 2019.
https://doi.org/10.1007/978-3-030-19153-5_46

The task scheduling of the earth observation satellites is a typical NP-Hard problem. In the process of multi-satellite coordinated observation, due to the different geographic locations and the different requirements in time delay and resolution of the observation tasks, as well as the limited time windows of the remote sensing satellites flying upon the observation tasks, the optimal matching method of satellite resources and tasks is difficult to find.

The existing scheduling models include integer linear programming model [4,6], backpack model [5], graph theory model [1,9,10], etc. The related algorithms are mostly heuristic algorithms such as simulated annealing algorithm and tabu search algorithm [9]. These algorithms have uncertain time complexity and cannot give approximate optimal solutions of NP-hard problems in polynomial time, resulting in high task acquisition delay and low utilization of satellite observation resources and energy resources.

In this paper, we will focus on the graph model representation of multi-satellite multi-task scenario in the earth observation process. Firstly, we will present the satellite observation scenario. Then, the weighted set cover modeling process will be introduced, followed by a greedy approximation algorithm, proved to be the optimal approximation algorithm in polynomial time for the minimum cost in our model.

2 WSC Model

2.1 Observation Scenario

Figure 1 shows the expansion diagram of multi-observation satellite flight around the earth over a period of time. The colored lines represent the satellite's flight paths, and the elliptic colored part represents the to-be-observed tasks.

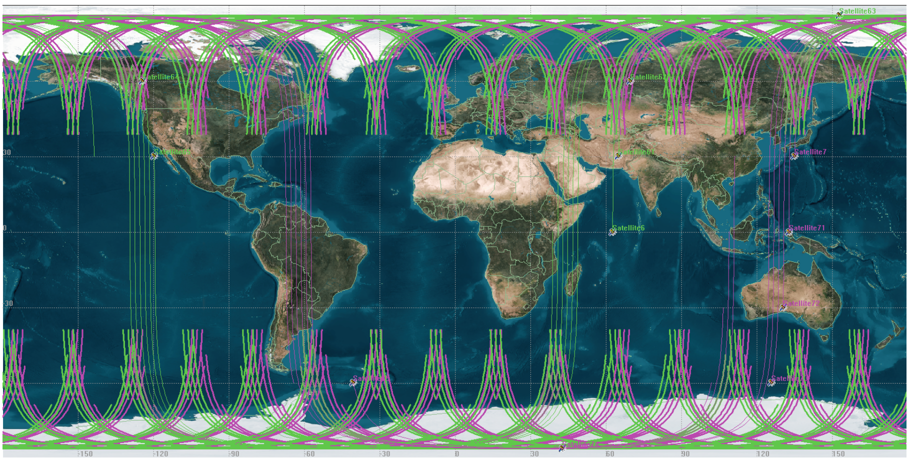


Fig. 1. The expansion diagram of multi-observation satellite flight around the earth over a period of time.

2.2 Modeling Process

Since the satellite trajectory is periodic, we consider the situation in which resources and tasks are predictable. In addition, due to the principle of fairness, the priority of observation tasks is not considered. Based on these principles, we model satellite observation scenarios.

Definition 1. *The parameters used in the following are defined as follows.*

- m_i*: The identification number of the original task,* $i = 1, 2, ...$*;*
- m_{ij}*: The identification number of the sub task split from* m_i*,* $j = 1, 2, ...$*;*
- $\mathcal{M}$*: The set of all sub tasks,* $\mathcal{M} = \{m_{ij}\,|i, j = 1, 2, ...\}$*;*
- T*: The total period of time considered;*
- τ_k*: The k-th discrete time slot;*
- sat_p*: The identification number of the satellite,* $p = 1, 2, ...$*;*
- $\mathcal{S}_{sat_p}^{\tau_k}$*: The collection set of the sub tasks which can be observed by* sat_p *within time slot* τ_k*;*
- $G_{\mathcal{S}_{sat_p}^{\tau_k}}$*: The gain of the set* $\mathcal{S}_{sat_p}^{\tau_k}$*, which is defined as the number of the elements in the set* $\mathcal{S}_{sat_p}^{\tau_k}$*;*
- $w_{\mathcal{S}_{sat_p}^{\tau_k}}$*: The wight of the set* $\mathcal{S}_{sat_p}^{\tau_k}$*, which is defined as the time window size required to observe all sub tasks in the set* $\mathcal{S}_{sat_p}^{\tau_k}$*;*

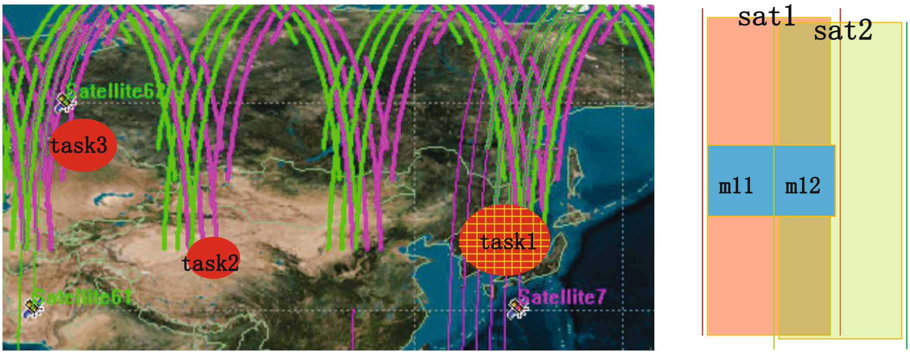
Firstly, the original tasks $m_i(i = 1, 2, ...)$ are split into sub tasks $m_{ij}(j = 1, 2, ...)$ of the same size according to experience or learning, satisfying $|m_i| = \bigcup_j |m_{ij}|$, where $|*|$ represents the geographic location of task $*$, as shown in Fig. 2(a) the task1.

Then, the total period of time we considered T is sliced into n time slots, donated in order as $\tau_1, \tau_2, ..., \tau_n$.

In each time slots, we model the collection of the sub tasks covered by each satellite as a set. The gain of each set can be expressed by the number of elements in the set, that is, the number of subtasks observed. In addition, a cost factor is attached to each set, defined here as the time required for all elements in the set to be covered, i.e. the length of time window required by the satellite to observe these subtasks.

For example, Fig. 2(b) shows the trajectory coverage of satellite sat_1 and satellite sat_2 in the same time slot τ_1, and the shaded parts represent subtasks m_{11} and m_{12}, then we get $\mathcal{S}_{sat_1}^{\tau_1} = \{m_{11}, m_{12}\}$, $G_{\mathcal{S}_{sat_1}^{\tau_1}} = 2$, $w_{\mathcal{S}_{sat_1}^{\tau_1}} = 1$, and $\mathcal{S}_{sat_2}^{\tau_1} = \{m_{12}\}$, $G_{\mathcal{S}_{sat_2}^{\tau_1}} = 1$, $w_{\mathcal{S}_{sat_2}^{\tau_1}} = 1$. That is, the two sets have the same cost, but the set $\mathcal{S}_{sat_1}^{\tau_1}$ covers more elements and the gain is larger, so the set $\mathcal{S}_{sat_1}^{\tau_1}$ is better than the set $\mathcal{S}_{sat_2}^{\tau_1}$. In addition, intuitively from Fig. 2 that storage resources will be wasted to observe non-target areas if satellite sat_2 were utilized in this area. The specific planning algorithm will be mentioned in Sect. 3.

According to the above modeling method, we construct the multi-satellite multi-task observation scenario as a weighted set cover model. Next we discuss the way of minimizing the cost on the premise of all tasks can be observed.



(a) (b)

Fig. 2. (a) Task splitting diagram (b) the example schematic diagram of flight paths of satellite sat_1 and satellite sat_2 in time slot τ_1

3 Approximate Greedy Algorithm

A large number of scholars have studied the degree of approximation of the set cover algorithm. The set cover was first proved to be NP-Complete by Karp [3], and Johnson [7] gave a polynomial time greedy approximation algorithm with an approximate ratio of $\ln n$. Chvatal [8] extended the algorithm to the weighted set cover problem. Feige et al. [2] proved that there is no polynomial time approximation algorithm better than approximation ratio $(1-\varepsilon)\ln X$, $(\varepsilon > 0)$, unless $NP \subseteq DTIME\left(n^{O(\log\log n)}\right)$. Zhang et al. [11] gave a greedy algorithm of ln $(|X|+1)$-approximation to the weighted set cover problem, according to Feige [2], this algorithm is optimal approximation algorithm in polynomial time.

Based on the proposed weighted set cover model, drawing on the classic greedy approximation algorithm, by adding the constraints of remote sensing satellite observation, an optimal approximation algorithm in polynomial time is given to optimize the time window of remote sensing satellite observation to save the remote sensing resources.

3.1 Constraints

The total on-off duration of each sensor on the satellite is limited by the satellite storage capacity, i.e., the formula below needs to be always established,

$$\sum_{k=1}^{n} \left\| \mathcal{S}_{sat_p}^{\tau_k} \right\| \leq C_{sat_p} \tag{1}$$

where $\left\| \mathcal{S}_{sat_p}^{\tau_k} \right\|$ represents the data sum of the subtask observation in the set, and C_{sat_p} represents the storage capacity of the satellite sat_p.

If the total data amount of the subtasks covered by satellite sat_p in all time slots exceeds the satellite storage capacity, the last overflow subtasks will be discarded according to the algorithm execution order.

3.2 Algorithm Flow

This paper draws on the thought of greedy algorithm proposed by Zhang [11], and considering the constraints of satellite scenario, the following approximate greedy algorithm suitable for the satellite scenario is given.

Algorithm 1. Approximate Greedy Algorithm

Input: $\mathcal{M}, \mathcal{S}_{sat_p}^{\tau_k}, w_{\mathcal{S}_{sat_p}^{\tau_k}}, C_{sat_p}(k = 1, 2, ..., n; p = 1, 2, ...)$
Output: a series of ordered sets $\mathcal{O} = \{\mathcal{O}_{sat_j}^{\tau_i} | i, j = 1, 2, ...\}$
$\mathcal{O} \leftarrow \phi$;
$\mathcal{V} \leftarrow \mathcal{M}$;
while $\mathcal{V} \neq \phi$ **do**
  Select the $\mathcal{S}_{sat_j}^{\tau_i}$ such that $\frac{w_{\mathcal{S}_{sat_j}^{\tau_i}}}{G_{\mathcal{S}_{sat_j}^{\tau_i} \cap \mathcal{V}}}$ is the smallest;
  if $\left\| \mathcal{O}_{sat_j} \cup \mathcal{S}_{sat_j}^{\tau_i} \right\| \leq C_{sat_j}$ **then**
    $\mathcal{V} = \mathcal{V} - \mathcal{S}_{sat_j}^{\tau_i}$;
    $\mathcal{O} = \mathcal{O} \cup \mathcal{S}_{sat_j}^{\tau_i}$;
  else
    $\mathcal{V} = \mathcal{V} - \sum_{i=1}^{n} \mathcal{S}_{sat_j}^{\tau_i}$;
  end if
  Output the set $\mathcal{O}$;
end while

The algorithm is a greedy algorithm. Each loop selects the smallest set of $\frac{w_{\mathcal{S}_{sat_j}^{\tau_i}}}{G_{\mathcal{S}_{sat_j}^{\tau_i} \cap \mathcal{V}}}$ from the remaining sets while ensuring that the observation data sum of each satellite does not exceed its total storage, until all subtasks are covered, where $G_{\mathcal{S}_{sat_j}^{\tau_i} \cap \mathcal{V}}$ represents the number of subtasks selected for the first time in the set $\mathcal{S}_{sat_j}$, and $w_{\mathcal{S}_{sat_j}^{\tau_i}}$ represents the cost (time window size) required to cover the set $\mathcal{S}_{sat_j}$. The algorithm returns the selected collection $\mathcal{O}$ in order of execution.

4 Approximate Analysis of the Greedy Algorithm

Theorem 1. *The approximation of the greedy algorithm proposed is lnn+O(1)-approximation of the optimal solution.*

Proof. We analyze the approximation of this algorithm briefly. In the greedy algorithm, the cost of selecting a set is $w_{\mathcal{S}^{\tau_i}_{sat_j}}$, and there are $G_{\mathcal{S}^{\tau_i}_{sat_j} \cap \mathcal{V}}$ subtasks that are covered for the first time in the set. So in this step, the cost of each node being covered is

$$cost(x_i) = \frac{w_{\mathcal{S}^{\tau_i}_{sat_j}}}{G_{\mathcal{S}^{\tau_i}_{sat_j} \cap \mathcal{V}}}. \tag{2}$$

Obviously, the total cost of all subtasks in the algorithm is

$$apx = \sum_i cost(x_i), \tag{3}$$

where *apx* is the cost of the greedy algorithm.

Now we consider the optimal solution. Suppose that at a certain step, the uncovered elements are $x_i, ..., x_n$, and the selected element is x_i, with another c_i elements selected, that is, the number of uncovered elements are at most c_i. The set cost is $w_{\mathcal{S}_i}$, so the following inequality holds,

$$opt \geq \frac{n-i+1}{c_i} * w_{\mathcal{S}_i}, \tag{4}$$

summing the left and right sides of the inequality respectively, we have

$$apx \leq opt * \sum_{i=1}^{n} \frac{1}{n-i+1}, \tag{5}$$

where $\sum_{i=1}^{n} \frac{1}{n-i+1} = \sum_{i=1}^{n} \frac{1}{i} = \ln n + c$, c is an Euler constant with a value of 0.6 approximately, i.e.,

$$apx \leq opt * (\ln n + O(1)). \tag{6}$$

According to Feige et al.'s proof at [2], the algorithm is the optimal polynomial time approximation algorithm for the weighted set cover model in the EOS scenarios.

5 Conclusion

In this paper, a weighted set cover model was proposed to characterize the matching of the tasks and the resource for the multi-satellite multi-task scenario. We put up one approximate greedy algorithm based on the model, where the constraint of satellite cache resources was considered, and the approximation of the algorithm to the optimal solution was analyzed.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (91638202,61871456,61401326,61571351), the National Key Research and Development Program of China (2016YFB0501004), the National S & T Major Project (2015ZX03002006), the 111 Project (B08038), Natural Science Basic Research Plan in Shaanxi Province of China (2016JQ6054).

References

1. Gabrel, V., Vanderpooten, D.: Enumeration and interactive selection of efficient paths in a multiple criteria graph for scheduling an earth observing satellite. Eur. J. Oper. Res. **139**, 533–542 (2002)
2. Feige, U.: A threshold of ln n for approximating set cover 1 introduction. J. ACM **45**(4), 314–318 (1999)
3. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W., Bohlinger, J.D. (eds.) Complexity of Computer Computations. The IBM Research Symposia Series, pp. 85–103. Springer, Boston (1972). https://doi.org/10.1007/978-1-4684-2001-2_9
4. Lin, W.C., Liao, D.Y., Liu, C.Y., Lee, Y.Y.: Daily imaging scheduling of an earth observation satellite. IEEE Trans. Syst. Man Cybern. - Part A: Syst. Hum. **35**(2), 213–223 (2005)
5. Shamna, T.P., Praveen, P.N.: Data acquisition and delay optimization in WSN using knapsack algorithm in presence of transfaulty nodes. In: 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1–5, June 2017
6. Sindhu, S., Sen, G.: An optimal scheduling policy for satellite constellation deployment. In: 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 100–104, December 2017
7. Johnson, D.S.: Approximation algorithms for combinatorial problems. J. Comput. Syst. Sci. **9**, 256–278 (1974)
8. Chvatal, V.: A greedy heuristic for the set-covering problem. Math. Oper. Res. **4**, 233–235 (1979)
9. Wu, G.H., Ma, M., Wang, H., Qiu, D.: Multi satellite observation scheduling based on task clustering. Acta Aeronautica et Astronautica Sinica **32**, 1275–1282 (2011)
10. Xu, Y.L., Xu, P.D., Wang, H.L., Peng, Y.H.: Clustering of imaging reconnaissance tasks based on clique partition. Oper. Res. Manag. Sci. **19**, 143–149 (2010)
11. Zhang, X.D., Luo, L.: Approximation algorithm for weighted set cover problem. J. Wenzhou Univ. Nat. Sci. **29**(6), 46–48 (2008)
12. Zhu, Y., Sheng, M., Li, J., Liu, R., Liu, J.: Modelling for data acquisition, storage and transmission of EOS. In: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1–6, October 2017