Journal of the Korean
OR/MS Society
Vol. 15, No. 2, Dec., 1990

Conditional Covering : Worst Case Analysis of Greedy Heuristics

I. Douglas Moon*

Abstract

The problem is a variation of the weighted set-covering problem (SCP) which requires the minimum-cost cover to be self-covering. It is shown that direct extensions of the well-known greedy heuristic for SCP can have an arbitrarily large error in the worst case. It remains an open question whether there exists a greedy heuristic with a finite error bound.

Keywords : Set Cover, Heuristics, Worst Case Analyses, Integer Program, Faculty Location.

1. Introduction

Imagine in the weighted set-covering problem (SCP) [4] that we require the ***facilities*** to cover not only all the ***customers*** but also one another. Such inter-facility covering is motivated by the desire to have one or more backup facilities to ensure a customer service level. Else, this secondary constraint is to enhance the facility-to-facility reinforcing capability as in some cases of distribution systems design. We shall call this variation of SCP the weighted ***conditional covering problem***(CCP). This study concerns the worst case performance of a class

* School of Management, The University of Michigan-Flint, Flint, Michigan

of greedy heuristics for CCP. In particular, we show that direct extensions of the well-known greedy heruistic for SCP (see[2] and [5]) can be arbitrarily bad. It remains an open question whether there exists a greedy type heuristic that gives a finite error. In this regard, we shall identify a specific research issue to be resolved in future studies.

Moon and Chaudhry [6] formulated CCP as an integer program for finding all x_j which will

$$\text{minimize} \sum_{j \varepsilon J} C_j X_j = Z$$

$$\text{subject to} \sum_{j \varepsilon J} a_{ij} x_j \geq 1 \qquad \text{for all } i \ \varepsilon \ I, \qquad (1)$$

$$\sum_{j \varepsilon J} b_{kj} x_j \geq x_k \qquad \text{for all } k \ \varepsilon \ J, \qquad (2)$$

$$x_j = (0, 1) \qquad \text{for all } j \ \varepsilon \ J, \qquad (3)$$

where $I = \{1, 2, \cdots, m\}$, the index set for customers,

$J = \{1, 2, \cdots, n\}$, the index set for potential facilities,

$$a_{ij} = \begin{cases} 1 \text{ if customer } i \text{ is covered by facility } j, \\ 0 \text{ otherwise,} \end{cases}$$

$$b_{kj} = \begin{cases} 1 \text{ if facility } k \text{ is covered by facility } j, \ k \neq j, \\ 0 \text{ otherwise,} \end{cases}$$

$c_j > 0$ for all j, the cost of facility j.

Constraint (1), as in SCP, ensures that each customer will be covered by at least one facility, whereas (2) is the additional constraint for covering each facility by at least one ***other*** facility. Inter-facility covering is not necessarily symmetric, in the sense that b_{kj} need not be the same as b_{jk}. Also note that the covering criterion for facilities can be different from that for customers.

In the absence of (2), the above formulation is the ordinary SCP. Since SCP is NP-Complete (see Garey and Johnson[3] for details), and it can be shown that CCP reduces to an extended SCP, CCP is also NP-Complete. Thus CCP can be considered intractably difficult to solve exactly for large problems.

To the best of our knowledge only unweighted CCP has been studied in the literature. Reference [6] used a linear programming relaxation method, and [1] presented extensive computational results of seven greedy heuristics.

We find it convenient to use the following notation in describing the heuristics discussed in this paper.

$P_j = \{i \in I | a_{ij} = 1\}$, $j \in J$, the set of customers covered by facility j,

$Q_j = \{k \in J | b_{kj} = 1\}$, $J \in J$, the set of facilities covered by facility j.

Using the sets (p_j) we see that the ordinary SCP is the problem of finding a cover $J^* \subseteq J$ which will satisfy $\cup(p_j : j \in J^*) = I$ at minimum $\Sigma(c_j : j \in J^*)$. Then CCP further stipulates that $\cup(Q_j : j \in J^*) \supseteq J^*$ as an additional constraint. We shall call the cover J^* that satisfies the latter condition a ***conditional*** cover.

For the remainder of this paper we assume (i) $j \notin Q_j$ for all j, (ii) $\cup \{P_j : j \in J\} = I$, and (iii) $\cup \{Q_j : j \in J\} = J$.

Before we present our heuristics for CCP in the next section, we give the set-covering heuristic and its worst error developed by Chvatal [2]. It is a greedy algorithm which selects one facility at a time based on the largest number of customers covered per unit cost, i. e., max $|p_j|/c_j$. The algorithm terminates as constraint (1) is met for all customers. Ho [5] asserted that the worst error of this heruistic is no greater than that of any other greedy type heuristic for SCP.

Set-Covering Heuristic (Chvatal)

Step 0. Set $J^* = \phi$.

Step 1. If $P_j = \phi$ for all j, then stop : J^* is a cover.

Otherwise find a subscript σ maximizing $|P_j|/c_j$ and proceed to Step 2.

Step 2. Add σ to J^*, replace each P_j by $P_j - P_\sigma$ and return to Step 1.

Theorem (Chvatal) The cost of the cover returned by the above heuristic is at most H(d) times the cost of an optimal cover where $H(d) = \Sigma(1/j : j = 1, 2, \cdots, d)$ with $d = \max|P_j|$.

2. Greedy Heuristics for CCP

We might expect that a natural extension of the set-covering heuristic would yield a comparable error ratio for CCP. We will consider two greedy heuristics. Heuristic 1 is a direct extension of the set-covering heuristic in the sense that it selects one facility at a time using $\max(|P_j| + |Q_j|)/c_j$ and that constraint (2) is part of stopping ciriterion as well. In contrast, Heuristic 2 uses the feasibility of constraint (2) as a prerequisite for facility selection. As shown next, both heuristics can perform poorly.

In Heuristic 1 we use the set J' to denote those facilities selected but not yet covered by any selected ones. Then we use the set R to denote *all* facilities, selected or not, which are covered by the selected facilities.

Heuristic 1

Step 0. Set $J^*=\phi$, $J'=\phi$, and $R=\phi$.

Step 1. If $P_j=\phi$ for all j and $J'=\phi$, then stop : J^* is a conditional cover.

Otherwise find a subscript σ maximizing $(|P_j|+|Q_j|)/c_j$ and proceed to Step 2.

Step 2. Add σ to J^* and to J' if $\sigma \notin R$. Add Q_σ to R and subtract Q_σ from J'. Proceed to Step 3.

Step 3. Replace every P_j by $P_j - P_\sigma$ and every Q_j by $Q_j - Q_\sigma$. Return to Step 1.

Proposition 1 : The worst case error for Heuristic 1 can be arbitrarily large.

Proof : Consider a problem having the parameters.

$J=\{1, 2, 3, 4\}$, $I=\{1, 2\}$

$P_1=\phi$, $P_2=\{1, 2\}$, $P_3=\{1, 2\}$, $P_4=\phi$,

$Q_1=\{2\}$, $Q_2=\{1\}$, $Q_3=\{4\}$, $Q_4=\{3\}$.

A graphical representation of these parameters is given in Figure 1 where an arc from j to i denotes i being covered by j for $j \in J$ and $i \in I, J$. If we assume $c_2<c_3<c_4<c_1$, then the heuristic finds a conditional cover $J^*=\{2, 3, 4, 1\}$, in that order of inclusion. If we assume $c_1+c_2>c_3+c_4$, then the optimal conditional cover is $\{3, 4\}$. Thus if Z_H is the cost of J^*, then

$$Z_H/Z=(c_1+c_2+c_3+c_4)/(c_3+c_4)=\frac{c_1+c_2}{c_3+c_4}+1.$$

If c_4 is a finite number, so are c_3 and c_2. But as $c_1\to\infty$, the ratio can be arbitrarily large.

Q. E. D.

It is worth noting that Heursitic 1 performs equally poorly even when performance measure $\max(|P_j|/c_j)$ or $\max(|Q_j|/c_j)$ is used. In the same example, either measure will initially select facility 2. Then, sooner or later, facility 1 must be selected, regardless of cost c_1,

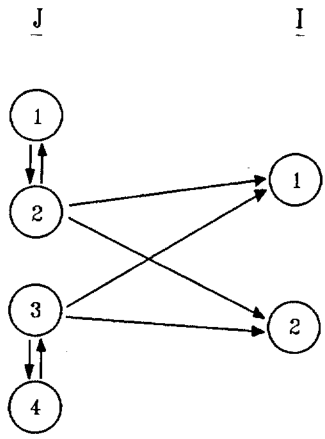

Figure 1

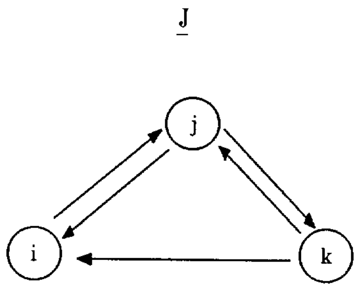

Figure 2 : Simple cycles

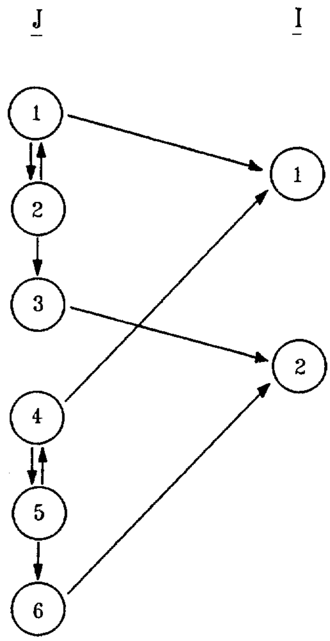

Figure 3

in order to cover facility 2.

The poor performance of Heuristic 1 may be due in part to (a) only one facility is selected at a time and (b) the feasibility of constraint (2) is part of stopping criteria. Heuristic 2 presented shortly relaxes both (a) and (b) by selecting one or **more** facilities at a time only when the feasibility of the associated constraint(s) (2) is guaranteed. Thus it will select a single facility from the set R described earlier or a set of facilities that can cover one another. We next explore the latter source of facilities to select.

Refer to Figure 2 where three facilities, i, j and k are shown. As before, the arcs represent the coverability : that is, $Q_i = \{j\}$, $Q_j = \{i, k\}$, and $Q_k = \{i, j\}$. We see that any directed cycle represents a self-covering set of facilities. In this example, cycles {i, j}, {j, k}, and {i, j, k} correspond to three (unordered) sets of self-covering facilities. The cycles of interest here, however, are simple ones. We say that a cycle is ***simple*** if no subset of the facilities in the cycle forms another cycle. In Figure 2, the cycles (i, j) and (j, k) are simple, but (i, j, k) is not. We note that generating all ***simple*** cycles based on constraint (2) is a polynomial time procedure.

We now give Heuristic 2 which selects a single facility from the set R or a set of facilities forming a simple cycle. We use the subscript π to denote such a facility or facilities. If π is not a single facility, then it is understood that $P\pi = \cup(P_j : j \in \pi)$, $Q\pi = \cup(Q_j : j \in \pi)$, and $c\pi = \Sigma(c_j : j \in \pi)$.

Heuristic 2

Step 0. Set $J^* = \phi$ and $R = \phi$.

Step 1. If $P_j = \phi$ for all j, then stop : J^* is a conditional cover. Otherwise find the subscript π such that $|P\pi|/c\pi$ is maximum, and proceed to Step 2.

Step 2. Add π to J^*, and $Q\pi$ to R. Proceed to Step 3.

Step 3. Replace every P_j by $P_j - P\pi$ and every Q_j by $Q_j - Q\pi$. Return to Step 1.

Note that the revision of Q_j in Step 3 is only to ensure R to contain nonredundant entries. Contrary to our expectation, the worst case performance of Heuristic 2 can be as poor as that of Heuristic 1. This is stated in Proposition 2 and proved by means of an example.

Proposition 2 : Heuristic 2 can be arbitrarily bad for CCP.

Proof : Our example has six facilities and two customers with the covering relationships given in Figure 3 and by

$$P_1 = P_4 = \{1\},\ P_3 = P_6 = \{2\},\ P_2 = P_5 = \phi,$$

$$Q_1 = \{2\},\ Q_2 = \{1, 3\},\ Q_3 = \phi,\ Q_4 = \{5\},\ Q_5 = \{4, 6\},\ Q_6 = \phi.$$

If $c_1 + c_2 + < c_4 + c_5$, then the heuristic selects the simple cycle {1, 2} over the simple cycle {4, 5} as the first self-covering set of facilities. Then facility 3, the only one in the set R, is selected over the cycle {4, 5}, whose $P\pi = \phi$, and the algorithm terminates. (Note that facility 6 has never been a candidate for selection.) If we assume $(c_1 + c_2) + c_3 > (c_4 + c_5) + c_6$, then the optimum conditional cover is {4, 5, 6}. Assuming that both $(c_1 + c_2)$ and $(c_4 + c_5)$ are arbitrarily small, we have

$$Z_H/Z = c/c_6.$$

While c_3 can be made arbitrarily large, c_6 can be arbitrarily small yielding an arbitrarily large error ratio.

Q. E. D.

3. Other Heuristics and Future Research

As we wish to cover all customers but only those facilities selected in CCP, a more natural extension of the set-covering heuristic would be one that uses $(|P_j|+|Q_j \cap J'|)$, toegether with the facility cost, as a performance measure. Regarding the facility cost, note that Heuristic 1 only considers the cost of the facility to be selected but ignores the cost of backup in case the facility is not in the set R. Thus, the following variation of facility cost is of interst :

$$c'_j = \begin{cases} c_j \text{ if } j \varepsilon \ R, \\ c_j + \min(c_k \mid j \varepsilon Q_k) \quad \text{otherwise.} \end{cases}$$

Based on the above arguments, one might feel that the following heuristic should be more promising, with intuitive appeal.

Heuristic 3 Same as Heuristic 1 but selects the next facility based on $\max(|P_j|+|Q_j \cap J'|)/c'_j$ in Step 1.

Reconsider the problem incidence used in the proof for Proposition 1. With those assumptions made about the facility costs, Heuristic 3 will find the optimal conditional cover {3, 4}. Unfortunately, there are other incidences for which Heuristic 3 dose not produce a finite error ratio. To illustrate, consider a problem having $J=\{0, 1, 2, 3, 4\}$, $I=\{1, 2\}$, with $P_0=\phi$, $P_1=\{2\}$, $P_2=\{1, 2\}$, $P_3=\{1, 2\}$, $P_4=\phi$, and $Q_0=\{1\}$, $Q_1=\{0, 2\}$, $Q_2=\phi$, $Q_3=\{4\}$, $Q_4=\{3\}$. If we assume c_0 is arbitrarily large and $c_1+c_2<c_3+c_4$, then Heuristic 3 finds a conditional cover {2, 1, 0}. But {3, 4} can be an optimal solution with a finite cost. Thus, Z_3/Z can again be arbitrarily large.

Supose we refine Heuristic 3 by restricting our attention to those facilities that can cover the hardest-to-cover customer. To this end, let $P_i^{-1}=\{j \varepsilon J | a_{ij}=1\}$ for each customer i not yet covered. Then let i* denote the customer with the smallest $|P_i^{-1}|$ among all customers yet to be covered. Now consider :

Heuristic 4 Same Heuristic 3 but evaluates only those facilities $j \in P_{i^*}^{-1}$ in Step 1. Using the problem incidence discussed above for Heuristic 3, we see that Heuristic 4 also fails to produce a finite error ratio.

We have considered several specific extensions of the set-covering heuristic for CCP. Each was shown to fail to guarantee a finite error ratio. The negative results like these can be more useful when they eliminate a whole class of heuristics from consideration, for they will save other research efforts. Thus, in concluding this paper, we identify the research issue to be resolved.

Let f be any function such that $f : R^+ \times Z^+ \rightarrow R$ where R^+ represents the set of costs like c'_j and Z^+ represents the set of integers like $(|P_j| + |Q_j \cap J'|)$. The object of future studies is to determine a value of r in the following statement.

For Future Research : Assume the greedy algorithm of Heuristic 3 or Heuristic 4 is used for CCP. There is no function f that gives a worst case error ratio strictly better than r where $r < \infty$.

If $r = \infty$ is established, then the greedy algorithms fail to yield a finite error ratio for CCP. Recall $r = H(d)$, developed by Chvatal, for the greedy heuristic for SCP[5].

References

[1] Chaudhry, S. S., I. D. Moon, and S. T. McCormick (1987). Conditional covering : Greedy Heuristics and Computational Results. *Computers and Opns. Res.* 14, pp. 11~18.

[2] Chvatal, V. (1979). A Greedy Heuristic for the Set-Covering Problem. *Math. Opns. Res.* 4, pp. 233~235.

[3] Garey, M. R. and D. S. Johnson (1979). *Computers and Intractability.* W. H. Freeman, San Francisco, CA.

[4] Garfinkel, R. S. and G. L. Nemhauser (1972). *Integer Programming.* John Wiley & Sons, New York.

[5] Ho, A. C. (1982). Worst Case Analysis of a Class of Set Covering Heuristics. *Math. Program.* 23, pp. 170~180.

[6] Moon, I. D. and S. S. Chaudhry (1984). An Analysis of Network Location Problems with Distance Constraints. *Management Science* 30, pp. 290~307.