rutcor amsmath,amssymb,euscript,graphicx 1 ex1 ex

A greedy heuristic for a generalized set covering problem

Peter ChenDepartment of Computer Science, Louisiana State University, Baton Rouge, Louisiana, 70803, USA (pchen@lsu.edu) Guoli DingDepartment of Mathematics, Louisiana State University, Baton Rouge, Louisiana, 70803, USA (ding@math.lsu.edu)

Peter Chen Guoli Ding

The authors are indebted to Robert Lax and Nigel Gwee for their valuable comments and suggestions. This research is partially supported by NSF grants DMS-0556091 and ITR-0326387.

The classical weighted set covering problem is generalized simultaneously in three directions. First, each numerical weight is replaced by a weighted set, which we call cost. Second, each element in the ground set is assigned a numerical weight. Third, the concept of a cover is relaxed to a partial cover that only needs to cover some percentage of the ground set, instead of the whole ground set. The last two generalizations have been studied in the literature, while the first is new. We propose a greedy algorithm to approximate this generalized problem and we establish an upper bound on the ratio of the greedy solution over the optimal solution. This bound is independent of the cost function, and it depends only on the total weight of the ground set. We prove that our bound is the best possible.

16-2008

document

Introduction

the purpose of this paper is to introduce a generalized set covering problem (GSCP), and to propose, in Section 2, a greedy algorithm to approximate it. We prove, in Theorem 1, that our solution is not too far away from the optimal solution. We also prove, in Theorem 2, that the bound given in Theorem 1 is the best possible.

For clarity, we first have some definitions. If X is a finite collection of sets, then $\overline{X}$ is the union of all members of X. If f is a function from a set X to R_+, the set of nonnegative reals, then, for any finite subset X' of X, $f(X')$ is defined to be the sum of $f(x)$, over all x in X'.

Let S be a finite set and let $S = \{S_1, S_2, \ldots, S_n\}$, where each S_i is a subset of S. We call $A \subseteq S$ a cover of S if $\overline{A} = \overline{S}$. The classical set covering problem (SCP) is to find a cover with a minimum cardinality. In applications, there is usually a weight function w from S to R_+. In such a situation, the total weight of a cover A is defined to be $w(A)$. The weighted SCP (WSCP) is to find a cover with a minimum total weight. Clearly, WSCP is a generalization of SCP, as WSCP is SCP when $w(S_i) = 1$, for all i.

In this paper, we study a generalization of WSCP, which arose from the authors' work on profiling. We will relax WSCP in three directions. First, we replace each numerical weight $w(S_i)$ with a weighted set. Second, we give every element in S a numerical weight. Third, we only require A to cover a portion of $\overline{S}$, instead of the entire $\overline{S}$.

Let S and S be as before. Let d be a function from S to R_+ and let $\lambda \in [0, 1]$. Then $A \subseteq S$ is called a λ-d-cover of S if $d(\overline{A}) \geq \lambda d(\overline{S})$. Notice that, when $d(x)$ is positive for all $x \in S$, then A is a cover if and only if it is a 1-d-cover. Therefore, "λ-d-cover" is a generalization of "cover".

Let W be a finite set, c be a function from W to R_+, and $W = \{W_1, W_2, \ldots, W_n\}$, where each W_i is a subset of W. We consider each W_i as the weight of S_i. For any $A \subseteq S$, we define $W(A) = \bigcup\{W_i : S_i \in A\}$ and we call $c(W(A))$ the cost of A. In particular, if $W = \{1, 2, ..., n\}$, and if for each i we have $W_i = \{i\}$ and $c(i) = w(S_i)$, then it is easy to see that $c(W(A))$ is exactly $w(A)$. Therefore, "cost" is a generalization of "total weight".

Generalized SCP (GSCP): For any given $(S, W, S, W, d, c, \lambda)$, find a λ-d-cover of S with the minimum cost.

In case $W = \{1, 2, ..., n\}$, $W = \{\{1\}, \{2\}, ..., \{n\}\}$, and $d(x) = 1$, for all $x \in S$, our GSCP is known as partial set cover problem (PSCP) kearns, which has the objective of finding $A \subseteq S$ with $|\overline{A}| \geq \lambda|\overline{S}|$ and such that $w(A)$ is minimized.

GSCP is also related to the submodular set cover problem (SSCP) wolsey, which we briefly describe below. Let U be a finite set and let f be a function from 2^U to the set of nonnegative integers such that: (i) $f(X) \leq f(Y)$ for all $X \subseteq Y \subseteq U$, and (ii) $f(X) + f(Y) \geq f(X \cap Y) + f(X \cup Y)$ for all $X, Y \subseteq U$. Let w be a function from U to R_+. The objective of SSCP is to find $A \subseteq U$ with $f(A) = f(U)$ and such that $w(A) = \sum_{a \in A} w(a)$ is minimized. It is not difficult to verify (see fujito) that SSCP is a special case of GSCP if $U = \mathcal{S}$ and $f(A) = \min\{\lambda d(\bar{S}), d(\bar{A})\}$, for all $A \subseteq S$, which is exactly the case of GSCP with $W = \{1, 2, ..., n\}$ and $W = \{\{1\}, \{2\}, ..., \{n\}\}$. Since the objective function of SSCP is linear while the objective function of GSCP is not, GSCP in general is not a special case of SSCP. On the other hand, since the constraint in SSCP is a general submodular function while the constraint in GSCP is a special submodular function (the function $f(A)$ defined above, see fujito), SSCP in general is not a special case of GSCP either. Therefore, GSCP and SSCP are incomparable, as illustrated in Figure 1.

figure[htbp] center [scale=1]comparison.eps GSCP and SSCP are different generalizations of PSCP.