INTERVAL SELECTION PROBLEMS WITH LIMITED OVERLAP

Andrew Ryzhikov, Mikhail Y. Kovalyov

United Institute of Informatics Problems, National Academy of Sciences, Belarus
6 Surganov str., 220072, Minsk, Belarus
kovalyov_my@yahoo.co.uk, ryzhikov.andrew@gmail.com

1. Introduction. Let $\mathcal{I}$ be a finite set of intervals on the real line. We assume that the endpoints of all intervals have rational coordinates. Every element $I \in \mathcal{I}$ has a non-negative rational weight $w(I)$. A set of intervals is called *independent* if no two intervals in this set have a common interior point, and *k-independent* if it is a union of k pairwise disjoint independent sets. Here k is a non-negative integer number. In the problem WEIGHTED k-INDEPENDENT SET OF INTERVALS the objective is to find a k-independent subset of $\mathcal{I}$ with maximum total weight. This problem is widely studied and has a lot of applications in interval scheduling, resource allocation, etc. For more details see surveys by Kovalyov et al. [1] and Kolen et al. [2]. It can be solved in polynomial time (Bouzina and Emmons [3]). We consider a generalization of this problem where the selected set of intervals must not be k-independent, but some overlap measure (which we call *composite u-redundancy*) of this set must be limited by a given number R. We define the composite u-redundancy in Section 2.

A special case of the WEIGHTED 1-INDEPENDENT SET OF INTERVALS problem is the problem where the weight of every interval is equal to its length. For arbitrary positive k we generalize this problem in the following way.There is a set $\mathcal{I}$ of intervals on the real line and the objective is to find a k-independent subset of $\mathcal{I}$ with maximum measure of the union. We call this problem MAXIMUM COVERAGE BY k-INDEPENDENT SET OF INTERVALS. We study this problem and also its generalization where the selected set must not be k-independent, but the composite u-redundancy of this set must be limited by a given number R.

2. Main definitions. Let u be a non-negative integer number and $\mathcal{J}$ be a subset of $\mathcal{I}$. We define the *set of projective u-redundancy* of $\mathcal{J}$ to be the set of such points on the real line that belong to at least $u+1$ intervals from $\mathcal{J}$. The measure of this set, that is, the total length of intervals in it, is called the *projective u-redundancy* of $\mathcal{J}$ and is denoted by $P(\mathcal{J}, u)$.

Further, let $x_1, \ldots, x_m$ be all the distinct left and right endpoints of intervals from $\mathcal{J}$ sorted in the increasing order. Let s_j be the number of intervals in $\mathcal{J}$ containing the interval $[x_j, x_{j+1}]$, $1 \le j \le m-1$. We define the *total u-redundancy* of the set $\mathcal{J}$ as

$$T(\mathcal{J}, u) = \sum_{j=1}^{m-1} \max\{(x_{j+1} - x_j) \cdot (s_j - u), 0\}.$$

Thus, in the projective u-redundancy, only one excessive interval of the intersection contributes to the redundancy value, and in the total u-redundancy all the excessive intervals of the intersection contribute to the redundancy value.

Let p and t be non-negative rational numbers such that $p + t > 0$. We define the *composite u-redundancy* of a set $\mathcal{J}$ to be the value $p \cdot P(\mathcal{J}, u) + t \cdot T(\mathcal{J}, u)$.

Both projective and total u-redundancy can be viewed as the measures that indicate the extent to which the set of intervals is not u-independent. In particular, the following lemma is true.

Lemma 1. *A finite set of intervals is k-independent if and only if its projective (or total) k-redundancy is equal to zero.*

3. Maximum weight selection problem. In the problem MAXWEIGHT we are given three integer numbers u, p, and t, an upper bound R on the composite u-redundancy and a ground set $\mathcal{I} = \{I_1, \ldots, I_n\}$ of intervals. Each interval is associated with a non-negative rational weight. The objective is to select a subset $\mathcal{J} \subseteq \mathcal{I}$ of the maximum total weight, provided that its composite u-redundancy does not exceed R. For $R = 0$ this problem is precisely the WEIGHTED u-INDEPENDENT SET OF INTERVALS problem.

The complexity of the MAXWEIGHT problem is characterized by the following theorem.

Theorem 1. *The* MAXWEIGHT *problem is NP-hard (in the ordinary sense) for any fixed u, p, t even if the weight of every interval is equal to its length and all endpoints of the intervals have integer coordinates.*

The next two theorems show that for any fixed u two restricted cases of the considered problem can be solved by pseudo-polynomial algorithms.

Theorem 2. *Let W be the total weight of all intervals in $\mathcal{I}$. There exists a pseudo-polynomial dynamic programming algorithm with running time $O(u^2Wn^{u+2})$ for the case where the weights of all intevals are integer numbers.*

Theorem 3. *There exists a pseudo-polynomial dynamic programming algorithm with running time $O(u^2(R+1)n^{u+2})$ for the case where the endpoints of all intervals have integer coordinates.*

It is an open question whether the general MAXWEIGHT problem is strongly NP-hard or pseudo-polynomially solvable.

4. Maximum coverage selection problem. The problem MAXCOVERAGE differs from the MAXWEIGHT problem in that the criterion is to maximize the measure of the union of the selected intervals, that is, the total length of the intervals of this union. For $R = 0$ this problem is precisely the MAXIMUM COVERAGE BY u-INDEPENDENT SET OF INTERVALS problem. The complexity of the MAXCOVERAGE problem is characterized by the following theorem.

Theorem 4. *The* MAXCOVERAGE *problem is NP-hard (in the ordinary sense) for arbitrary fixed non-negative rational numbers p and t, and for both $u = 0$ and $u = 1$.*

One special case of this problem can be solved in pseudo-polynomial time.

Theorem 5. *Let L be the union measure of all intervals in $\mathcal{I}$. There exists a pseudo-polynomial dynamic programming algorithm with running time $O(Ln^{u+2})$ for the case where $u \in \{0, 1\}$ and the endpoints of all intervals have integer coordinates.*

Theorem 6. *There exists a $\frac{1}{2}$-approximation algorithm with running time $O(n \log n)$ for the* MAXCOVERAGE *problem with $u = 1$.*

We also prove that for the developed algorithm the number $\frac{1}{2}$ in this bound cannot be replaced with a larger constant.

According to the following theorem, the case $u \geq 2$ is much simpler.

Theorem 7. *There exists an algorithm with running time $O(n \log n)$ that finds a 2-independent subset $\mathcal{J}$ of intervals such that the union of the intervals in $\mathcal{J}$ coincides with the union of the intervals in $\mathcal{I}$.*

This implies that the problems MAXCOVERAGE and MAXIMUM COVERAGE BY u-INDEPENDENT SET OF INTERVALS are solvable in $O(n \log n)$ time for $u \geq 2$.

The work is partially supported by the Belarusian Republican Foundation for Fundamental Research under the grant number Ф15СО-043.

References

1. Kovalyov M.Y., Ng C.T., Cheng T.C.E. *Fixed interval scheduling: Models, applications, computational complexity and algorithms* // European Journal of Operational Research. 2007. V. 178. No. 2. P. 331–342.

2. Kolen A.W.J., Lenstra J.K., Papadimitriou C.H., Spieksma F.C.R. *Interval scheduling: A survey* // Naval Research Logistics. 2007. V. 54. No. 5. P. 530–543.

3. Bouzina K.I., Emmons H. *Interval Scheduling on identical machines* // Journal of Global Optimization. 1996. V. 9. No. 3–4. P. 379–393.