Об одном подходе к решению задачи о покрытии с интервальными весами и его вычислительной сложности

А. В. Пролубников

Омский государственный университет им. Ф.М. Достоевского, Россия

Контактный e-mail: a.v.prolubnikov@mail.ru

Рассмотрена задача о покрытии множества с интервальными весами подмножеств. Решение задачи основано на получении множества приближенных решений для всех возможных весов из заданных интервалов — объединенного приближенного множества решений задачи. Такой подход использует модификацию жадного алгоритма для решения задачи о покрытии на случай интервальных весов. Количество приближенных решений, содержащихся в объединенном приближенном решении, определяет вычислительную сложность его нахождения. Показано, что зависимость вычислительной сложности предложенного подхода от радиусов интервалов весов множеств является неубывающей кусочно-постоянной функцией. Приведен пример приложения подхода к решению задачи формирования рейсов для сети железных дорог при неопределенностях в стоимостях обслуживания рейсов. В заключение работы ставится задача коррекции численно неразрешимых задач о покрытии с интервальными весами.

Ключевые слова: задача о покрытии множества, интервальная неопределенность, вычислительная сложность.

Введение

Взвешенная задача о покрытии множества (далее обозначаемая аббревиатурой ЗПМ) имеет следующую постановку. Даны конечное множество U мощности m и конечный набор множеств $S = \{ S_1, \dots, S_n\}$ такой, что $S_i \subseteq U$ и $\cup_{i=1}^n S_i = U$. Набор множеств $S' = \{ S_{i_1}, \dots, S_{i_k}\}$, $S_{i_j} \in S$, называется *покрытием* множества U, если $\cup_{i=1}^k S_{i_j} = U$. Для $S_i \in S$ заданы *веса* $w_i = w\,(S_i)$, $w_i > 0$. Вес $w\,(S')$ набора множеств S' равен сумме весов входящих в него множеств. Необходимо найти покрытие множества U минимального веса — *оптимальное покрытие*.

Задача о покрытии множеств NP-трудна [1], и ее часто требуется решать в качестве вспомогательной постановки при решении других оптимизационных задач и задач анализа данных. Приложения задачи о покрытии множеств включают в себя задачи размещения [2], машинного обучения [3], распределения ресурсов [4], интеллектуального анализа данных [5].

В рассматриваемой нами интервальной постановке этой задачи функция веса w на множествах из S принимает интервальные значения, т. е. для каждого множества из S заданы минимальный и максимальный возможные веса. Неопределенность веса множества в приложениях может быть вызвана неопределенностью во входных данных, например погрешностью измерений или вариабельностью соответствующих величин.

В настоящей работе предлагается подход к решению задачи о покрытии с интервальными весами (которую далее будем называть ИЗПМ) и исследуется вычислительная сложность решения ИЗПМ с помощью представленного в статье подхода.

1. Задача о покрытии с интервальными весами множеств и жадный алгоритм ее решения

Задача о покрытии с интервальными весами. Для обозначения интервалов будем использовать, в соответствии с неформальным международным стандартом, жирный курсивный шрифт. Для интервала $\boldsymbol{a} \subset \mathbb{R}$ его нижняя и верхняя границы обозначаются соответственно как $\underline{\boldsymbol{a}}$ и $\overline{\boldsymbol{a}}$. Декартово произведение интервалов $\boldsymbol{a}_1 \cdot \ldots \cdot \boldsymbol{a}_n$ представляет собой *брус* в $\mathbb{R}^n$ или, иными словами, интервальный вектор.

В задаче о покрытии с интервальными весами $\mathcal{P}$ заданы множества U, S и вектор $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ интервальных весов множеств из S, $\underline{\boldsymbol{w}}_i > 0$.

Определение 1. *Реализацией весов множеств для ИЗПМ $\mathcal{P}$ называется вектор точечных весов $w \in \boldsymbol{w}$.*

Пусть $\mathcal{P}_w$ обозначает ЗПМ, получаемую из $\mathcal{P}$ при замене в ней интервальных весов $\boldsymbol{w}$ на реализацию $w \in \boldsymbol{w}$. Для различных w как оптимальные, так и полученные с помощью какого-либо алгоритма приближенные решения $\mathcal{P}_w$ могут различаться.

Определение 2. *Приближенным решением $\mathcal{P}$ будем называть покрытие для $\mathcal{P}_w$ для некоторой реализации $w \in \boldsymbol{w}$, полученное с помощью некоторого алгоритма, понимаемое как упорядоченный набор множеств.*

Определение 3. *Объединенное оптимальное множество решений $\mathcal{P}$ с заданным брусом возможных весов $\boldsymbol{w}$ — это множество оптимальных решений, содержащее хотя бы одно оптимальное решение $\mathcal{P}_w$ для любой реализации $w \in \boldsymbol{w}$.*

Определение 4. *Объединенное приближенное множество решений $\mathcal{P}$ с заданным брусом возможных весов $\boldsymbol{w}$ — это множество приближенных решений, содержащее хотя бы одно приближенное решение $\mathcal{P}_w$ для любой реализации $w \in \boldsymbol{w}$.*

Объединенное приближенное множество решений ИЗПМ $\mathcal{P}$ будем обозначать посредством $\mathrm{Gr}[\mathcal{P}]$, так что $\mathrm{Gr}[\mathcal{P}] = \{Gr_1, \ldots, Gr_N\}$ — конечное множество покрытий, где $N = |\mathrm{Gr}[\mathcal{P}]|$, а Gr_i — приближенные решения задач $\mathcal{P}_w$ для $w \in \boldsymbol{w}$.

Нами ставится задача нахождения приближенного решения ИЗПМ как задача полного или частичного получения ее объединенного приближенного множества решений. Везде далее в качестве приближенных решений ИЗПМ будем рассматривать только решения (покрытия), получаемые с помощью точечного жадного алгоритма для возможных реализаций весов множеств.

Интервальный жадный алгоритм. Напомним схему жадного алгоритма решения ЗПМ, модификацией которого является интервальный жадный алгоритм [6]. В ходе работы жадного алгоритма решения ЗПМ SCP множества из S выбираются в покрытие Gr соответственно минимуму *относительных весов* $w_i/|S_i|$ множеств $S_i \in S$ до тех пор, пока набранные в Gr множества не образуют покрытия множества U.

Замечание 1. Жадный алгоритм является детерминированным, т. е. задано некоторое правило выбора на шаге 4 множества S_q в том случае, если вариантов выбора несколько. Если после выполнения шага 7 получаем, что $S_i = \varnothing$, то S_i исключается из S.

Жадный алгоритм решения ЗПМ $(SCP) : Gr$;

$Gr \leftarrow \varnothing$;
while $U \neq \varnothing$
    begin
        выбрать $S_q \in S$ такое, что $w_q/|S_q| = \min\left\{ w_i/|S_i| \,\middle|\, S_i \in S \text{ и } S_i \not\subseteq \bigcup_{j \in Gr} S_j \right\}$;
        $Gr \leftarrow Gr \cup \{S_q\}$;
        $U \leftarrow U \setminus S_q$;
        for $i \leftarrow 1$ **to** $i = n$
            $S_i \leftarrow S_i \setminus S_q$;
    end;
print Gr;

При небольшой трудоемкости, составляющей $O(m^2 n)$, для общего случая ЗПМ с положительными весами [7] верна логарифмическая оценка точности веса приближенного решения, получаемого с помощью жадного алгоритма:

$$\frac{w(Gr)}{w(Opt)} \leq H(m) \leq \ln m + 1, \tag{1}$$

где $H(m) = \sum\limits_{k=1}^{m} 1/k$; Gr — покрытие, получаемое с помощью жадного алгоритма; Opt — оптимальное покрытие. В [8] показано, что при условии $P \neq NP$ получение полиномиального алгоритма с оценкой точности $(1-\varepsilon)\ln m$ при $\varepsilon > 0$ невозможно, т. е. жадный алгоритм является асимптотически лучшим полиномиальным алгоритмом для ЗПМ. Представленную выше схему жадного алгоритма назовем *точечным* жадным алгоритмом.

Будем использовать следующие обозначения. Для ИЗПМ $\mathcal{P}$ покрываемое множество обозначим как $U[\mathcal{P}]$, заданный набор его подмножеств — как $S[\mathcal{P}]$, вектор интервальных весов этих множеств — как $\boldsymbol{w}[\mathcal{P}]$.

В ходе работы *интервального жадного алгоритма* решения ИЗПМ [6], оперирующего с интервальными относительными весами множеств, производится перебор множеств допустимых реализаций весов, дающих возможные варианты построения покрытия точечным жадным алгоритмом, и производится построение этих покрытий.

Интервальный жадный алгоритм в качестве входа получает $\mathcal{P}$ и производит вызов процедуры, реализующей схему рекурсии с возвратом. В ходе рекурсивных вызовов производится перебор возможных вариантов построения приближенных решений $\mathcal{P}$.

Для ИЗПМ $\mathcal{P}$ определим $Q[\mathcal{P}]$ как множество допустимых, исходя из возможных относительных весов, вариантов выбора на первой итерации решения $\mathcal{P}$ жадным алгоритмом: $Q[\mathcal{P}] = \{S_{i_1}, \ldots, S_{i_q}\}$, где $S_{i_j} \in Q[\mathcal{P}]$, если существует такая реализация $w \in \boldsymbol{w}[\mathcal{P}]$, что

$$w_{i_j}/|S_{i_j}| = \min\left\{ w_i/|S_i| \;\middle|\; S_i \in S[\mathcal{P}] \right\}. \tag{2}$$

Работа точечного жадного алгоритма может быть представлена как последовательное построение (на отдельных итерациях) задач о покрытии, каждая из которых получается из предыдущей при выборе в покрытие множества с минимальным относительным весом. При этом из множества S предыдущей задачи удаляется выбранное в покрытие множество, а из U и множеств из S удаляются элементы, покрытые выбранным множеством. Аналогично работа интервального жадного алгоритма при решении ИЗПМ $\mathcal{P}$ представляет собой построение последовательностей ИЗПМ $\mathcal{P}^{(i_1,\ldots,i_k)}$ (рис. 1), где $\mathcal{P}^{(0)}=\mathcal{P}$, а задача $\mathcal{P}^{(i_1,\ldots,i_k)}$ получается из задачи $\mathcal{P}^{(i_1,\ldots,i_{k-1})}$ такой же корректировкой множеств U, S и их элементов, что и выше в схеме жадного алгоритма при выборе в Gr множества S_{i_k}. Брус $\boldsymbol{w}[\mathcal{P}^{(i_1,\ldots,i_k)}]$ получается из $\boldsymbol{w}[\mathcal{P}^{(i_1,\ldots,i_{k-1})}]$ удалением реализаций, несовместимых с достижением на S_{i_k} минимума относительных весов множеств из $S[\mathcal{P}^{(i_1,\ldots,i_{k-1})}]$.

Таким образом, в ходе работы интервального алгоритма для ИЗПМ $\mathcal{P}$ получаем дерево задач $T_{\mathcal{P}}$. В нем две ИЗПМ из множества его вершин $V(T_{\mathcal{P}})$ смежны, если ИЗПМ, соответствующая вершине-потомку, получается из ИЗПМ, соответствующей вершине-предку после вышеописанной модификации.

Если на брусе $\boldsymbol{w}[\mathcal{P}]$ задано вероятностное распределение, то с помощью интервального жадного алгорима могут быть рассчитаны вероятности приближенных решений из $\mathrm{Gr}[\mathcal{P}]$.

Определение 5. *Вероятностью* $\mathsf{P}(Gr)$ *приближенного решения* Gr *для* $\mathcal{P}$ *назовем вероятность получения такой реализации* $w\in\boldsymbol{w}$*, что* Gr *— решение* $\mathcal{P}_w$*.*

Имея на итерации интервального жадного алгоритма задачу $\mathcal{P}'\in V(T_{\mathcal{P}})$, для каждого множества из $Q[\mathcal{P}']$ рассчитываем вероятность получения такой реализации $w\in\boldsymbol{w}[\mathcal{P}']$, что для $\mathcal{P}'$ минимум в (2) будет достигаться на этом множестве, т. е. для $S_i\in Q[\mathcal{P}']$ рассчитываются вероятности $\mathsf{P}(S_i)$ их выбора в строящееся на итерациях алгоритма покрытие $Gr\in\mathrm{Gr}[\mathcal{P}]$, $\sum\limits_{S_i\in Q[\mathcal{P}']}\mathsf{P}(S_i)=1$. Для построенного $Gr=\{S_{i_1},\ldots,S_{i_k}\}\in\mathrm{Gr}[\mathcal{P}]$ имеем

$$\mathsf{P}(Gr)=\prod_{j=1}^{k}\mathsf{P}_{Gr}(S_{i_j}),$$

где $\mathsf{P}_{Gr}(S_{i_j})$ — значение, рассчитанное при включении S_{i_j} в Gr. Поскольку $\mathrm{Gr}[\mathcal{P}]$ содержит решения для всех реализаций $w\in\boldsymbol{w}[\mathcal{P}]$, то $\sum\limits_{Gr\in\mathrm{Gr}[\mathcal{P}]}\mathsf{P}(Gr)=1$. Имея распределение на брусе $\boldsymbol{w}[\mathcal{P}]$, с помощью интервального жадного алгоритма получаем распределение на множестве приближенных решений из $\mathrm{Gr}[\mathcal{P}]$.

В результате работы интервального жадного алгоритма для ИЗПМ $\mathcal{P}$ получаем:

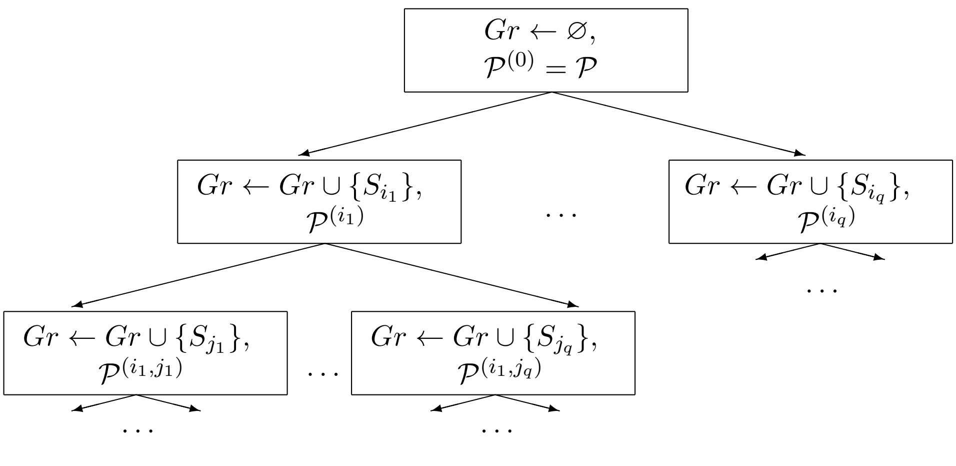

Рис. 1

1) объединенное приближенное множество решений $\mathrm{Gr}[\mathcal{P}]$;
2) брусы $\boldsymbol{w}_{Gr} \subseteq \boldsymbol{w}[\mathcal{P}]$ для $Gr \in \mathrm{Gr}[\mathcal{P}]$ — внешние оценки подмножеств таких реализаций $w \in \boldsymbol{w}[\mathcal{P}]$, что Gr — решение ЗПМ $\mathcal{P}_w$;
3) получаемые с помощью оценок $\boldsymbol{w}_{Gr} \in \mathbb{IR}^n$ оценки $\boldsymbol{w}(Gr) \in \mathbb{IR}$ веса покрытия Gr:

$$\boldsymbol{w}(Gr) = \sum_{S_i \in Gr} (\boldsymbol{w}_{Gr})_i,$$

где $(\boldsymbol{w}_{Gr})_i$ — i-я компонента бруса $\boldsymbol{w}_{Gr}$;
4) вероятности $\mathsf{P}(Gr)$ приближенных решений $Gr \in \mathrm{Gr}[\mathcal{P}]$.

Для всех приближенных решений из $\mathrm{Gr}[\mathcal{P}]$ справедлива логарифмическая оценка гарантированной точности.

2. Подход к решению ИЗПМ, использующий интервальный жадный алгоритм

Представляемый подход к решению ИЗПМ позволяет с помощью интервального жадного алгоритма получать объединенное приближенное множество решений ИЗПМ и производить его анализ с целью оценивания возможных значений целевой функции для приближенных решений. В результате применения этого подхода, как указано выше, можно получить объединенное приближенное множество решений, интервальные оценки весов входящих в него покрытий, а также, как будет показано далее, при заданном вероятностном распределении на интервалах весов может быть определено вероятностное распределение веса приближенного решения на множестве возможных значений целевой функции. Понимая значение целевой функции как величину затрат, можно использовать эту информацию в ситуации неопределенности весов для оценивания возможных затрат.

Пример использования подхода. Рассмотрим задачу о покрытии со следующей содержательной постановкой. Пусть задан граф железных дорог, ребра которого соответствуют железнодорожным перегонам, а вершины — станциям. Указано, на каких станциях располагаются депо. Требуется, чтобы количество перегонов в маршруте не превышало двух и чтобы маршрут начинался и заканчивался в одном и том же депо, проходя по составляющим его перегонам в обоих направлениях. Заданы веса возможных маршрутов — стоимости обслуживания рейсов по этим маршрутам, включающие заработную плату машинистов, стоимость обслуживания подвижного состава и др. Необходимо составить такой набор маршрутов, чтобы все перегоны были ими покрыты и затраты на обслуживание рейсов по этим маршрутам были минимальными.

Пусть граф задачи — это граф с выделенными вершинами, соответствующими депо (рис. 2). Возможные маршруты из S заданы перечислением входящих в них перегонов

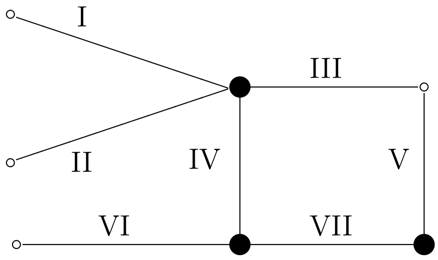

Рис. 2

в порядке их следования по мере удаления от депо: $S_1 = \{\text{I}\}$, $S_2 = \{\text{II}\}$, $S_3 = \{\text{III}\}$, $S_4 = \{\text{IV}\}$, $S_5 = \{\text{VI}\}$, $S_6 = \{\text{VII}\}$, $S_7 = \{\text{V}\}$, $S_8 = \{\text{III,V}\}$, $S_9 = \{\text{IV,VI}\}$, $S_{10} = \{\text{IV,VII}\}$, $S_{11} = \{\text{VII,VI}\}$, $S_{12} = \{\text{IV,I}\}$ $S_{13} = \{\text{IV,II}\}$, $S_{14} =\{\text{VII,V}\}$. Стоимости обслуживания рейсов по этим маршрутам составляют: $w_1 = 80$, $w_2 = 100$, $w_3 = 120$, $w_4 = 90$, $w_5 = 100$, $w_6 = 130$, $w_7 = 130$, $w_8 = 200$, $w_9 = 210$, $w_{10} = 200$, $w_{11} = 210$, $w_{12} = 210$, $w_{13} = 230$, $w_{14} = 250$.

Интервальная неопределенность в стоимостях обслуживания рейсов, являющихся в этой постановке весами множеств ЗПМ, может возникать, например, в связи с льготами и премиальными вознаграждениями машинистов, региональными коэффициентами к зарплате и др. Допустим, что для некоторых трех выделенных маршрутов из числа заданных набором множеств S стоимости обслуживания рейсов задаются не точно, а как, например, интервалы вида $[w_i - 40, w_i + 40]$. В этом случае в зависимости от выбора таких выделенных маршрутов возможно получение ИЗПМ, для которых количество возможных приближенных решений будет сильно различаться.

Так, в случае, если интервальными становятся, например, веса множеств S_2, S_3, S_5, получаем 82 решения в объединенном приближенном множестве решений при максимальной вероятности решения, равной 0.096, средней — 0.012, при среднеквадратичном отклонении 0.0002. Полученные интервалы значений целевой функции для приближенных решений из объединенного приближенного множества решений позволяют оценить возможный интервал ее значений как $[580, 785]$.

Если интервальными становятся веса множеств S_1, S_{12}, S_{14}, то получаем три решения, различающиеся только порядком входящих в них множеств: $\mathrm{Gr} = \{Gr_1, Gr_2, Gr_3\}$, где $Gr_1 = \{S_1, S_4, S_2, S_5, S_8, S_6\}$, $Gr_2 = \{S_4, S_1, S_2, S_5, S_8, S_6\}$, $Gr_3 = \{S_4, S_2, S_5, S_8, S_1, S_6\}$ с вероятностями их получения, соответственно равными 0.625, 0.125 и 0.250. Интервалы значений целевой функции составляют соответственно $[660, 710]$, $[710, 720]$ и $[720, 740]$. Исходя из этого можно сделать вывод о том, что приближенное решение, которое будет получено точечным жадным алгоритмом, одно и то же для всех возможных реализаций весов, тогда как интервал возможных значений затрат составит $[660, 740]$.

Получение вероятностного распределения на множестве возможных значений целевой функции. Предположим, что в ходе работы интервального жадного алгоритма для ИЗПМ получено множество $\mathbf{w}(\mathrm{Gr}) \subseteq \mathbb{IR}$ возможных значений целевой функции для приближенных решений из Gr, представляющее собой следующее объединение интервалов:

$$\mathbf{w}(\mathrm{Gr}) = \bigcup_{Gr \in \mathrm{Gr}} \boldsymbol{w}(Gr) = \bigcup_{Gr \in \mathrm{Gr}} \left(\sum_{S_i \in Gr} (\boldsymbol{w}_{Gr})_i \right).$$

Замечание 2. Множество $\mathbf{w}(\mathrm{Gr})$ возможных значений целевой функции для приближенных решений из Gr обозначается не так, как в работе обозначаются интервалы, поскольку оно может представлять собой как один интервал, так и несвязное объединение интервалов (мультиинтервал).

Имея вероятности $\mathsf{P}(Gr)$ для $Gr \in \mathrm{Gr}$ и интервалы возможных весов $\boldsymbol{w}(Gr)$, определяемые по $\boldsymbol{w}_{Gr}$, мы можем определить распределение вероятности веса приближенного решения на $\mathbf{w}(\mathrm{Gr})$. Для функции плотности $\mathsf{p}(w)$ этого распределения имеем

$$\mathsf{p}(w) = \sum_{Gr \in \mathrm{Gr}} \mathsf{p}(w|Gr)\mathsf{P}(Gr),$$

где $\mathsf{p}(w|Gr)$ — плотность распределения веса в w при условии получения реализации, для которой точечным жадным алгоритмом будет получено Gr. Если на интервалах весов задано равномерное распределение, то с ростом количества множеств в покрытии плотность $\mathsf{p}(w|Gr)$, согласно центральной предельной теореме теории вероятностей, стремится к нормальному распределению с плотностью

$$\mathsf{p}(w|Gr_i) = \frac{1}{\widehat{\sigma}\sqrt{2\pi}} \exp\left(-\frac{(w-\widehat{w})^2}{2\widehat{\sigma}^2}\right), \tag{3}$$

где $\widehat{w} = \sum\limits_{S_j \in Gr} w_j$, $\widehat{\sigma}^2 = \sum\limits_{S_j \in Gr} \sigma_j^2$ для средних значений w_j и средних квадратичных отклонений σ_j равномерно распределенного на $\boldsymbol{w}_j$ веса множества S_j: $w_j = (\underline{\boldsymbol{w}}_j + \overline{\boldsymbol{w}}_j)/2$, $\sigma_j = \sqrt{(\overline{\boldsymbol{w}}_j - \underline{\boldsymbol{w}}_j)^2/12}$. По результатам имитационного моделирования функция распределения вероятности веса целевой функции на интервале $\boldsymbol{w}(Gr)$ пренебрежимо мало отличается от (3) уже при количестве слагаемых в $\boldsymbol{w}(Gr) = \sum\limits_{S_j \in Gr} (\boldsymbol{w}_{Gr})_j$, превышающем 10. Для небольших значений $|Gr|$ могут быть использованы формулы для распределения вероятности суммы независимых случайных весов, получаемые через формулу для вычисления свертки плотностей вероятностей слагаемых, аналогично тому, как это делается в [9].

3. Зависимость вычислительной сложности применения предложенного подхода от величины радиусов интервалов весов

Вычислительная сложность применения подхода. Заданный брус $\boldsymbol{w}[\mathcal{P}]$ и комбинаторная структура $\mathcal{P}$ определяют дерево $T_{\mathcal{P}}$, мощность множества его вершин, а значит, и вычислительную сложность решения индивидуальной ИЗПМ $\mathcal{P}$ интервальным жадным алгоритмом и вычислительную сложность предложенного подхода в целом. Общая вычислительная сложность подхода складывается из следующих составляющих:

1) вычислительной сложности интервального жадного алгоритма, включая нахождение $\mathrm{Gr}[\mathcal{P}]$, $\boldsymbol{w}_{Gr} \in \mathbb{IR}^n$ и $\boldsymbol{w}(Gr) \in \mathbb{IR}$ для всех $Gr \in \mathrm{Gr}$;
2) вычислительной сложности нахождения вероятностей $\mathsf{P}(Gr)$ покрытий $Gr \in \mathrm{Gr}$, если заданы распределения вероятностей весов на $\boldsymbol{w}[\mathcal{P}]$;
3) вычислительной сложности нахождения функции плотности распределения вероятности веса на $\mathbf{w}(\mathrm{Gr})$.

При этом для ИЗПМ $\mathcal{P}' \in V(T_{\mathcal{P}})$ множество $Q[\mathcal{P}']$ и соотношение концов интервалов относительных весов входящих в него множеств могут быть таковы, что вычислительная сложность нахождения вероятностей выбора множеств в покрытие, а значит и сложность вычисления вероятностей покрытий и нахождения распределения на множестве значений $\mathbf{w}(\mathrm{Gr}[\mathcal{P}])$, будут экспоненциально расти с увеличением m и n [6].

Вычислительная сложность $C_1(\mathcal{P})$ нахождения $\mathrm{Gr}[\mathcal{P}]$, оценок множеств реализаций $\boldsymbol{w}_{Gr}$ для $Gr \in \mathrm{Gr}[\mathcal{P}]$ и их вероятностей равна общей сложности всех $N(T_{\mathcal{P}})$ итераций интервального жадного алгоритма, где $N(T_{\mathcal{P}}) = |V(T_{\mathcal{P}})|$:

$$C_1(\mathcal{P}) = O(mn)N(T_{\mathcal{P}}) + \sum_{\mathcal{P}' \in V(T_{\mathcal{P}})} C_2(\mathcal{P}'). \tag{4}$$

Здесь $O(mn)$ — вычислительная сложность одной итерации интервального жадного алгоритма; $C_2(\mathcal{P}')$ — вычислительная сложность процедуры вычисления вероятностей выбора множеств из $Q[\mathcal{P}']$ в покрытие Gr. Если определить значение $C_2(\mathcal{P})$ как

$$C_2(\mathcal{P}) = \max_{\mathcal{P}' \in V(T_{\mathcal{P}})} \Big\{ C_2(\mathcal{P}') \Big\},$$

то, как следует из (4), верна оценка

$$C_1(\mathcal{P}) \le (O(mn) + C_2(\mathcal{P}))N(T_{\mathcal{P}}). \tag{5}$$

Вычислительная сложность $C_3(\mathcal{P})$ расчета значения $\mathsf{p}(w)$ в соответствии с (3) составляет $O(m|\mathrm{Gr}[\mathcal{P}]|)$, так как в каждом покрытии из $\mathrm{Gr}[\mathcal{P}]$ содержится не более m множеств. Поскольку максимальная длина ориентированной цепи в дереве $T_{\mathcal{P}}$ не превышает m, то $N(T_{\mathcal{P}}) \le m|\mathrm{Gr}[\mathcal{P}]|$ и общая вычислительная сложность $C(\mathcal{P}) = C_1(\mathcal{P}) + C_3(\mathcal{P})$ применения подхода к решению ИЗПМ $\mathcal{P}$ по (5) может быть оценена как $C(\mathcal{P}) \le (O(m^2 n) + mC_2(\mathcal{P}))|\mathrm{Gr}[\mathcal{P}]|$. Поскольку $O(mn)N(T_{\mathcal{P}}) \le C(\mathcal{P})$, то

$$O(mn)|\mathrm{Gr}[\mathcal{P}]| \le C(\mathcal{P}) \le (O(m^2 n) + mC_2(\mathcal{P}))|\mathrm{Gr}[\mathcal{P}]|, \tag{6}$$

где константы в $O(mn)$ и $O(m^2 n)$ фиксированы. Из этого заключаем, что значение $|\mathrm{Gr}[\mathcal{P}]|$ может служить параметром, характеризующим вычислительную сложность ИЗПМ $\mathcal{P}$.

Зависимость вычислительной сложности подхода от радиусов интервалов весов. Далее, исследуя вычислительную сложность решения ИЗПМ, будем считать, что для ИЗПМ $\mathcal{P}(\delta)$ радиусы интервалов весов множеств из $S[\mathcal{P}(\delta)]$ равны δ, т. е. компоненты вектора интервалов весов $\boldsymbol{w}[\mathcal{P}(\delta)]$ имеют вид $\boldsymbol{w}_i(\delta) = [w_i - \delta, w_i + \delta]$. Без ограничения общности относительно возможных линейных зависимостей можно считать, что линейные по δ зависимости $\underline{\boldsymbol{w}}_i(\delta)$ и $\overline{\boldsymbol{w}}_i(\delta)$ имеют указанный вид, поскольку принадлежность $Q[\mathcal{P}(\delta)]$ определяется относительными весами множеств, которые при такой зависимости могут принимать любые возможные значения на итерациях интервального жадного алгоритма. С ростом δ мощность $Q[\mathcal{P}(\delta)]$ может увеличиваться, как это показано на рис. 3. Так, на рис. 3, *а*, где $\delta = 0$, имеем $|Q[\mathcal{P}(\delta)]| = 1$, тогда как при значениях δ, не равных нулю, для рис. 3, *б* и *в* $|Q[\mathcal{P}(\delta)]| = 2$ и $|Q[\mathcal{P}(\delta)]| = 3$ соответственно. Это означает увеличение $|\mathrm{Gr}[\mathcal{P}(\delta)]|$.

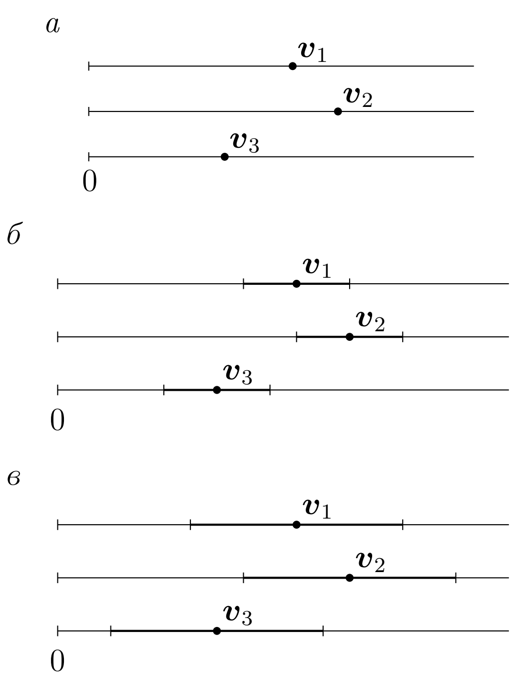

Рис. 3

Установим вид зависимости мощности множества $\mathrm{Gr}[\mathcal{P}(\delta)]$ от δ, что будет означать и вид функции $C(\mathcal{P}(\delta))$ вычислительной сложности решения ИЗПМ $\mathcal{P}(\delta)$ с помощью предложенного подхода. Для интервального веса $\boldsymbol{w}_i(\delta) = [w_i - \delta, w_i + \delta]$ и $p_i = |S_i|$ имеем интервал относительных весов $\boldsymbol{v}_i(\delta)$:

$$\boldsymbol{v}_i(\delta) = \left[\frac{w_i}{p_i} - \frac{\delta}{p_i}, \frac{w_i}{p_i} + \frac{\delta}{p_i}\right].$$

Соотношение концов интервалов относительных весов множеств из $S[\mathcal{P}(\delta)]$ определяет набор элементов множества $Q[\mathcal{P}(\delta)]$. По мере изменения δ в точках пересечения прямых, являющихся графиками $\underline{\boldsymbol{v}}_i(\delta)$ и $\overline{\boldsymbol{v}}_i(\delta)$, может меняться взаимный порядок относительных весов множеств из $Q[\mathcal{P}(\delta)]$, а в некоторых из них (на рис. 4 это точки с абсциссами δ_1 и δ_2) мощность множества $Q[\mathcal{P}(\delta)]$ может увеличиваться. В этом случае будет увеличиваться и количество концевых вершин в дереве $T_{\mathcal{P}}$, равное $|\mathrm{Gr}[\mathcal{P}(\delta)]|$. При линейном по δ изменении границ интервалов весов зависимость $|\mathrm{Gr}[\mathcal{P}(\delta)]|$ является неубывающей, поскольку по определению множества $Q[\mathcal{P}(\delta)]$ принадлежность $S_i \in Q[\mathcal{P}(\delta)]$ эквивалентна

$$\underline{\boldsymbol{v}}_i(\delta) \leq \min_{S_j \in S[\mathcal{P}(\delta)]} \overline{\boldsymbol{v}}_j(\delta).$$

Кроме того, если это верно для некоторого δ_0, то верно и для любого $\delta > \delta_0$. На рис. 4 изображены линейные зависимости границ интервалов ИЗПМ, где $p_1 = p_3$.

Таким образом, для ИЗПМ $\mathcal{P}(\delta)$ зависимость $|\mathrm{Gr}[\mathcal{P}(\delta)]|$ от величины радиусов интервалов весов δ является неубывающей кусочно-постоянной функцией (рис. 5). В точках δ_i на рис. 5 меняется мощность множеств $Q[\mathcal{P}'(\delta)]$ для $\mathcal{P}'(\delta) \in V(T_{\mathcal{P}(\delta)})$. Значение $|\mathrm{Gr}[\mathcal{P}(0)]|$ равно количеству вариантов построения покрытия интервальным жадным алгоритмом, когда на его итерациях при выборе минимального значения на шаге 4 встречаются множества с одинаковыми точечными относительными весами.

Замечание 3. Кусочно-постоянная неубывающая функция $|\mathrm{Gr}[\mathcal{P}(\delta)]|$ является ограниченной, поскольку при ограниченном количестве точек пересечения прямых $\underline{\boldsymbol{v}}_i(\delta)$ и $\overline{\boldsymbol{v}}_i(\delta)$, $i = \overline{1, n}$, имеется ограниченное количество точек δ_i, в которых увеличивается мощность $Q[\mathcal{P}'(\delta)]$ для какой-либо $\mathcal{P}'(\delta) \in V(T_{\mathcal{P}(\delta)})$. Следовательно, существует такое δ_r, что для любого $\delta > \delta_r$ справедливо $|\mathrm{Gr}[\mathcal{P}(\delta)]| = |\mathrm{Gr}[\mathcal{P}(\delta_r)]|$.

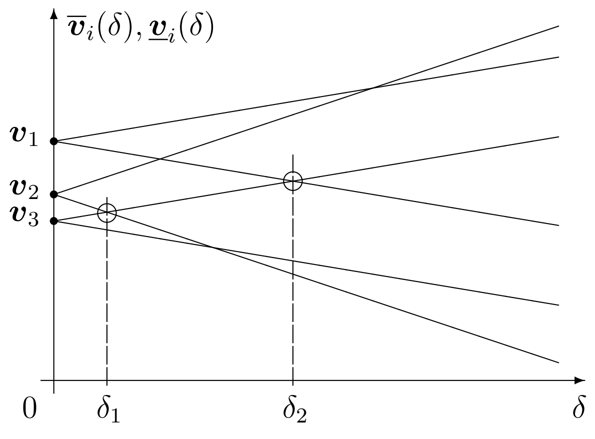

Рис. 4

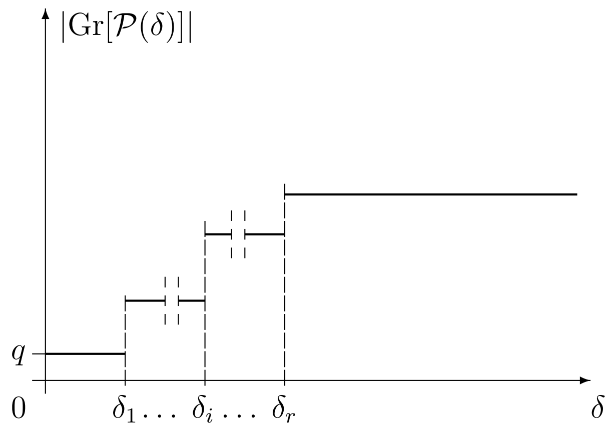

Рис. 5

4. Вычислительно неразрешимые ИЗПМ и их коррекция

Будем считать, что для ИЗПМ $\mathcal{P}$ на брусе $\boldsymbol{w}[\mathcal{P}] \in \mathbb{IR}^n$ задано некоторое вероятностное распределение.

Предложенный в работе подход может быть применен к индивидуальной ИЗПМ $\mathcal{P}$ в том случае, если вычислительная сложность интервального жадного алгоритма для $\mathcal{P}$, которая, вообще говоря, экспоненциальна, будет относительно невелика. При этом для $\mathcal{P}$ могут быть численно получены множества $\mathrm{Gr}[\mathcal{P}]$, а также получена вся сопутствующая информация о решениях, ему принадлежащих. Но нетрудно представить ситуацию, когда мощность $\mathrm{Gr}[\mathcal{P}]$ окажется настолько велика, что с учетом оценки (6) применение предложенного в этой работе подхода будет невозможным ввиду слишком большой вычислительной сложности $C(\mathcal{P})$.

Интервальный жадный алгоритм может быть модифицирован [6] так, что для $\mathcal{P}$ в качестве результата он будет давать не $\mathrm{Gr}[\mathcal{P}]$, а $\mathrm{Gr}_\varepsilon[\mathcal{P}]$, определяемое как

$$\mathrm{Gr}_\varepsilon[\mathcal{P}] = \{Gr \in \mathrm{Gr}[\mathcal{P}] : \mathsf{P}(Gr) \geq \varepsilon\}.$$

Параметр ε — это минимальная вероятность решений из $\mathrm{Gr}[\mathcal{P}]$, при которой они могут быть учтены при оценивании множества возможных значений целевой функции $\mathbf{w}(\mathrm{Gr}_\varepsilon[\mathcal{P}])$ для решений из $\mathrm{Gr}_\varepsilon[\mathcal{P}]$ и построении на нем распределения. Значение

$$\mathsf{P}(\mathrm{Gr}_\varepsilon[\mathcal{P}]) = \sum_{Gr \in \mathrm{Gr}_\varepsilon[\mathcal{P}]} \mathsf{P}(Gr_i)$$

— это вероятность получения такой реализации $w \in \boldsymbol{w}[\mathcal{P}]$, что приближенное решение ЗПМ $\mathcal{P}_w$ будет содержаться в $\mathrm{Gr}_\varepsilon[\mathcal{P}]$.

Определение 6. *Будем говорить, что ИЗПМ $\mathcal{P}$ ε-неразрешима, если $\mathrm{Gr}[\mathcal{P}]$ не содержит решений с вероятностью, большей ε. В противном случае будем говорить, что ИЗПМ ε-разрешима.*

ε-неразрешимость ИЗПМ можно трактовать как неустойчивость ЗПМ к внесению в нее неопределенности. Значение $\mathsf{P}(\mathrm{Gr}_\varepsilon[\mathcal{P}])$ может служить мерой разрешимости ИЗПМ $\mathcal{P}$ с помощью предложенного подхода. $\mathsf{P}(\mathrm{Gr}_{\varepsilon_1}[\mathcal{P}]) \geq \mathsf{P}(\mathrm{Gr}_{\varepsilon_2}[\mathcal{P}])$ при $\varepsilon_1 < \varepsilon_2$, т. е. рассмотрение большего количества приближенных решений дает бóльшую достоверность получаемых оценок. Для ε-неразрешимой ИЗПМ $\mathcal{P}$ имеем $\mathsf{P}(\mathrm{Gr}_\varepsilon[\mathcal{P}]) = 0$.

Уже при относительно небольших размерностях ИЗПМ может иметь большую вычислительную сложность и быть ε-неразрешимой для малых значений ε. Пусть, к примеру, $\varepsilon = 0.001$ и пусть $\alpha \in [0, 1]$ — доля множеств, веса которых являются невырожденными интервалами. Для сгенерированных нами задач о вершинном покрытии для графов на 100 вершинах ($n = 100$) и с 200 ребрами ($m = 200$) ε-неразрешимые ИЗПМ были получены уже при радиусах интервалов весов $\delta = 3$ и значении $\alpha = 0.2$. При $\alpha = 0.5$ и том же δ такие ИЗПМ составили уже около 50 % от всех сгенерированных задач, при $\alpha = 0.8$ — около 80 %. Возможные значения центров интервальных весов в этих задачах составляли $10, \ldots, 15$. Для задач больших размерностей, представленных в [10], где $m = 200$, $n \geq 1000$ и центры весов устанавливались равными $10, \ldots, 110$, ε-неразрешимые задачи были получены уже при $\alpha = 0.02$ и том же значении δ, а при $\alpha = 0.1$ практически любая из сгенерированных ИЗПМ оказывалась ε-неразрешимой.

В связи с этим имеет смысл поставить *задачу коррекции ИЗПМ*, т. е. задачу получения ε-разрешимой ИЗПМ из ε-неразрешимой ИЗПМ. Производимая при этом модификация интервальных весов множеств должна быть оптимальной относительно некоторого критерия, который может пониматься, например, как минимизация дополнительных затрат на уточнение интервальных оценок весов множеств, достаточное для получения численно разрешимой ИЗПМ. Другими критериями, исходя из которых может решаться задача коррекции для ε-неразрешимой задачи, могут быть, к примеру, следующие:

1) минимизация суммарных изменений интервалов весов, при которых ИЗПМ становится ε-разрешимой;
2) нахождение такого наименьшего по мощности набора множеств из S с невырожденными интервальными весами, что после замены весов этих множеств на центры их интервальных весов ИЗПМ становится ε-разрешимой.

Для коррекции ИЗПМ $\mathcal{P}(\delta)$ может быть использована информация о кусочно-постоянной неубывающей зависимости вычислительной сложности решения ИЗПМ от радиусов интервалов весов при их линейном росте. При этом мы рассчитываем значения δ_i, при которых происходит изменение мощности $\mathrm{Gr}[\mathcal{P}(\delta)]$, а значит, и вычислительной сложности применения предложенного подхода к решению $\mathcal{P}(\delta)$, и выбираем такое наибольшее значение δ, при котором $\mathcal{P}(\delta)$ ε-разрешима.

Выводы

В работе представлен подход к решению задачи о покрытии множества с интервальными весами, в ходе которого строится объединенное приближенное множество решений, получается множество возможных значений целевой функции, и, если задано вероятностное распределение на брусе допустимых весов множеств, могут быть получены вероятности принадлежащих ему приближенных решений и вероятностное распределение на нем. Показано, что вычислительная сложность нахождения решения задачи о покрытии с интервальными весами с помощью предложенного подхода имеет кусочно-постоянную, не убывающую от радиусов интервалов весов зависимость. Дана постановка задачи коррекции для вычислительно сложных индивидуальных задач о покрытии множества с интервальными весами.

Список литературы / References

[1] **Гэри М., Джонсон Д.** Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. 416 с.
Garey, M., Johnson, D. Computers and intractability: a guide to the theory of NP-completeness. New York: W.H. Freeman & Company, 1978. 352 p.

[2] **Farahani, R.Z., Asgari, N., Heidari, N., Hosseininia, M., Goh, M.** Survey: covering problems in facility location: A review // Computers and Industrial Engineering. 2012. Vol. 62, No. 1. P. 368–407.

[3] **Kearns, M.J.** The computational complexity of machine learning. Cambridge: The MIT Press, 1990. 176 p.

[4] **Milena, M.** Set cover with requirements and costs evolving over time // Randomization, Approximation, and Combinatorial Optimization. Algorithms and Techniques: Lecture Notes in Computer Science. Berlin; Heidelberg: Springer, 1999. Vol. 1671. P. 63–72.

[5] **Gao, B.J., Ester, M., Cai, J.-Y., Schulte, O., Xiong, H.** The minimum consistent subset cover problem and its applications in data mining // Proc. of the 13th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, KDD '07. New York, USA, 2007. P. 310–319.

[6] **Пролубников А.В.** Задача о покрытии множества с интервальными весами подмножеств и жадный алгоритм ее решения // Вычисл. технологии. 2015. Т. 20, № 6. С. 70–84.
Prolubnikov, A.V. The set-cover problem with interval weights and the greedy algorithm for its solution // Comput. Technologies. 2015. Vol. 20, No. 6. P. 70–84. (In Russ.)

[7] **Chvatal, V.** A greedy heuristic for the set-covering problem // Mathematics of Operation Research. 1979. Vol. 4, No. 3. P. 233–235.

[8] **Feige, U.** A threshhold of $\ln n$ for approximating set cover // J. of the ACM (JACM). 1998. Vol. 45, No. 4. P. 634–652.

[9] **Шарый С.П.** Интервальный анализ или методы Монте-Карло? // Вычисл. технологии. 2007. Т. 12, № 1. С. 103–115.
Shary, S.P. Interval analysis or Monte-Carlo methods? // Comput. Technologies. 2007. Vol. 12, No. 1. P. 103–115. (In Russ.)

[10] OR-Library. Available at: http://people.brunel.ac.uk/~mastjjb/jeb/info.html (accessed: 25.04.2016).

Поступила в редакцию 14 февраля 2017 г.

On an approach to set covering problem with interval weights and its computational complexity

Prolubnikov, Alexander V.

Omsk F.M. Dostoevsky State University, Omsk, 644077, Russia

Corresponding author: Prolubnikov, Alexander V., e-mail: a.v.prolubnikov@mail.ru

Purpose. To obtain an approximation of the optimal solution of the set covering problem with interval weights and to estimate weights of the possible solutions. To estimate computational complexity of numerical realization of the presented approach.

Methodology. Discrete optimization, interval analysis, probability theory.

Findings. An approach to obtain the estimations of weights for the possible solutions of the set covering problem with interval uncertainties is presented. We show that complexity of the numerical realization of the approach is non-decreasing step function of the interval widths. It is shown that the considered problem may be computationally hard even for small values of the parameters. An approach to the modification of such problems by means of their correction is presented. It facilitates the reduction of the problem to make it close to the initial one in accordance with certain criterion and keeps it nevertheless numerically solvable

Originality/value. An approach for obtaining an approximate solution of the covering problem with interval uncertainties in the weights is presented. Estimates of the weights for possible solutions of the problem is obtained. The computational complexity of the numerical implementation of the approach is estimated. It is shown that the covering problem with interval weights can be computationally complex for small values of its parameters.

Keywords: the set covering problem, interval uncertainty, computational complexity.

Received 14 February 2017