УДК 519.87

О представительности приближённых решений задач дискретной оптимизации с интервальной целевой функцией

А.В. Пролубников
Омский государственный университет
e-mail: a.v.prolubnikov@mail.ru

Рассматриваются задачи дискретной оптимизации с интервально заданной неопределённостью, моделирующей ошибку измерений значений коэффициентов целевой функции. Под возможным оптимальным решением задачи понимается допустимое решение, которое при некоторых возможных значениях коэффициентов, принимающих значения из заданных интервалов, является оптимальным. Вероятность получения возможного оптимального решения — это вероятность получения таких значений коэффициентов, при которых это решение будет оптимальным. Под возможным приближённым решением понимается решение, полученное с помощью определённого приближённого алгоритма решения задачи для некоторых возможных значений коэффициентов. Вероятность получения такого решения определяется как вероятность получения значений коэффициентов, при которых оно будет получено алгоритмом. В том случае, если вероятность получения возможного (оптимального или приближённого) решения меньше заданного порогового значения, решение считается нами непредставительным. Среднее (оптимальное или приближённое) решение — это решение, получаемое для средних значений из интервалов возможных значений коэффициентов. Мы показываем, что доля индивидуальных задач с интервальной неопределённостью коэффициентов, в которых как среднее, так и любое другое приближённое решение непредставительно, может быть достаточно велика при малой относительной погрешности измерений и пороге представительности близком к нулю.

Ключевые слова: дискретная оптимизация, интервальная неопределённость, приближённые решения.

1. Введение

Конечной целью прикладных научных исследований является надёжный прогноз результатов предстоящих экспериментов. Это касается и прикладных исследований в рамках такой математической дисциплины, как дискретная оптимизация. Алгоритмы, реализующие методы дискретной оптимизации, используют в качестве входных данных результаты измерений параметров некоторой математической модели, которая устанавливает соотношения между совокупностью переменных задачи. Эта математическая модель включает в себя *целевую функцию*, задающую такую числовую характеристику моделируемого явления, что большему (для задачи на максимум) или меньшему (для задачи на минимум) её значению будет соответствовать лучшая с точки зрения лица, принимающего решение, ситуация в реальной жизни. Предпочтительность возможных ситуаций может характеризоваться значением потерь (стоимостью затрат), которые будут понесены в случае принятия того или иного решения.

Рассматриваемые нами задачи дискретной оптимизации (далее — *задачи ДО*) имеют следующую постановку.

Постановка задачи (I). Пусть $E = \{e_1, \ldots, e_n\}$. $c(e) > 0$ — это *стоимость* элемента $e \in E$, $c_i = c(e_i)$. Двоичный вектор $x = (x_1, \ldots, x_n)$ задаёт множество $E_x \subset E$: $x_i = 1$, если $e_i \in E_x$, и $x_i = 0$, если $e_i \in E \setminus E_x$. Задано *множество допустимых решений* $\mathcal{D}$. Необходимо найти *оптимальное решение* $\check{x} \in \mathcal{D}$, дающее минимум целевой функции

$$f(x, c) = \sum_{e \in E_x} c(e) = \sum_{i=1}^{n} c_i x_i. \quad (1)$$

Так могут быть поставлены многие задачи ДО на графах и гиперграфах. А именно, множество E может представлять собой множество рёбер графа, тогда как множество допустимых решений $\mathcal{D}$ может быть рассмотрено как множество двоичных векторов, ассоциированных с подграфами заданного вида в нём — с множеством путей, соединяющих вершины в графе, с множеством остовных деревьев, с множеством гамильтоновых циклов, с паросочетаниями, разрезами и т.д. Задачи ДО в такой постановке не ограничиваются задачами на графах (гиперграфах). Например, так может быть сформулирована задача о булевом рюкзаке. Различные прикладные задачи могут быть поставлены как задачи ДО такого вида: задачи управления запасами, задачи размещения, задачи календарного планирования и др.

На практике точные измерения параметров задач ДО, как правило, невозможны. Неопределённость в значениях входных данных может присутствовать и по причине неполноты информации на момент начала исследований и возможности её изменений с течением времени. Часто единственной информацией относительно неопределёного параметра может быть только то, что его значение принадлежит некоторому интервалу на числовой оси. Рассматривая коэффициенты целевой функции как значения из заданных интервалов, будем считать, что на них задано наименее информативное вероятностное распределение — равномерное.

Реализацией будем называть вектор значений коэффициентов целевой функции, взятых по одному из заданного для каждого из них интервала. Под *возможным оптимальным решением* задачи ДО с интервально заданной неопределённостью коэффициентов (далее — *задача ИДО*) будем понимать допустимое решение, которое для некоторой реализации является оптимальным. Аналогично, *возможным α-приближённым решением* задачи ИДО называется решение $\tilde{x} \in \mathcal{D}$ такое, что при некоторых возможных значениях c имеем $f(\tilde{x}, c) \leq \alpha f(\check{x}, c)$, $\alpha \geq 1$. Далее будем считать, что значение α фиксировано, и будем называть такое решение *возможным приближённым решением*.

Оптимальное решение, получаемое для средних значений из заданных для кооэффициентов интервалов, будем называть *средним оптимальным решением* и обозначать как $\check{x}_\mu$. Возможное приближённое решение для средних значений коэффициентов будем называть *средним приближённым решением* и обозначать как $\tilde{x}_\mu$.

Центр интервала, которому принадлежат результаты многократных измерений параметра, часто используется исследователями как оценка неизвестного точного значения параметра. Это может приводить к тому, что, получив только одно возможное решение — среднее оптимальное или среднее приближённое, соответствующее усреднённым значениям неизвестных параметров, мы игнорируем другие возможные решения задачи в ситуации неопределённости параметров. Далее мы покажем, что довольно часто такой подход может оказываться необоснованным.

При заданном вероятностном распределении на интервалах возможных значений коэффициентов под *вероятностью получения возможного оптимального решения* нами понимается вероятность получения такой реализации, что данное допустимое решение будет оптимальным. Под *вероятностью получения возможного приближённого решения* для некоторого приближённого алгоритма решения задачи мы понимаем вероятность получения такой реализации, при которой это решение будет получено алгоритмом. В том случае, если вероятность получения возможного (оптимального или приближённого) решения меньше заданного порогового значения b — *порога представительности*, решение считается нами *непредставительным*, в противном случае — *представительным.*

Для NP-трудных задач ДО нахождение оптимального решения за приемлемое время в общем случае не может быть обеспечено уже для задач с несколькими сотнями переменных. Разработаны алгоритмы, дающие для таких задач за полиномиальное время приближённые решения с гарантированной точностью. В качестве примера такого алгоритма нами будет рассматриваться жадный алгоритм для задачи о покрытии множества.

Как при поиске оптимальных решений, так и при приближённом решении задач с неточно заданными параметрами возникают следующие вопросы, касающиеся неопределённости ситуации, получаемой при переходе от работы с точными значениями параметров к работе с интервально заданными параметрами.

1. Насколько много возможных (оптимальных или приближённых) решений будет иметь задача ДО с интервальной неопределённостью параметров?
2. Насколько сильно будут отличаться значения целевой функции для них?

Поиски ответов на эти и схожие с ними вопросы в различных постановках производились с тех пор, как стали разрабатываться методы решения задач ДО с точными значениями параметров. Как правило, эти вопросы формулировались с использованием концепции *устойчивости* оптимального решения задачи ДО, под которой понималась неизменность оптимального решения задачи ДО при варьировании значений коэффициентов целевой функции в заданных интервалах.

Предлагаемая нами концепция *представительности* возможных решений задачи ДО близка по своему смыслу к концепции устойчивости, однако принципиальным является то, что с её помощью мы пытаемся оценить, насколько много может быть возможных решений при допущении неопределённости параметров задачи и насколько велика вероятность получить то же самое решение, что и в случае их определённости. Возможное приближённое решение может быть устойчивым, то есть для него может существовать $\delta > 0$ такое, что в δ-окрестности $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$ оно будет единственным приближённым решением, получаемым с помощью жадного приближённого алгоритма, однако, при этом оно не будет представительным для некоторых значений δ и b. При заданных интервальных ограничениях на коэффициенты мы можем иметь такое множество возможных оптимальных или приближённых решений, что и среднее, а, как правило, и любое другое возможное приближённое решение не будут иметь достаточной вероятности получения для того, чтобы считаться представительными.

Таким образом, любое возможное оптимальное (или приближённое) решение рассматривается нами как один из представителей всего множества возможных оптимальных (или приближённых) решений. Если задача ДО с интервальной целевой функцией моделирует повторяющуюся ситуацию, что отражается заданием вероятностного распределения на множестве реализаций, то возможное решение тем представительней, чем чаще мы его будем получать с помощью некоторого используемого нами алгоритма. Если моделируется ситуация, в которой выбор некоторого решения из множества допустимых необходимо произвести только один раз, то решение будет тем представительней, чем больше будет вероятность получения такой реализации, при которой оно будет оптимальным или α-приближённым.

Кроме того, введение такой характеристики решения, как «представительность», нам удобно, поскольку даёт краткость формулировок. Так, например, формулировка «решение имеет вероятность получения меньшую некоторого заданного порогового значения» при заданном пороге представительности будет сокращаться нами до формулировки «решение непредставительно», а формулировка «доля задач, в которых вероятность получения среднего решения не меньше заданного порогового значения» — до формулировки «доля задач, в которых среднее решение непредставительно».

В работе показывается, что как средние приближённые, так и любые другие возможные приближённые решения задачи ДО в постановке (I) могут оказываться непредставительным при относительно небольших погрешностях задания коэффициентов целевой функции (1) и небольших значениях порога представительности. При росте размерности задачи непредставительными могут оказываться все возможные приближённые решения. Поэтому для прогнозирования потерь в ситуации неопределённости нам может потребоваться поиск всех возможных приближённых решений задачи и анализ таких их характеристик, как вероятности получения, интервалы возможных значений целевой функции для них, вероятностное распределение на нём и др.

2. Задачи дискретной оптимизации с интервальной целевой функцией

Интервальное представление неопределённости. Аналогично тому, как это делается в [2] и во многих других публикациях по интервальному анализу, интервальные значения далее будем обозначать жирным шрифтом:

$$\boldsymbol{a} = [\underline{\boldsymbol{a}}, \overline{\boldsymbol{a}}] = \{a \in \mathbb{R} \mid \underline{\boldsymbol{a}} \leq a \leq \overline{\boldsymbol{a}}\},$$

где $\underline{\boldsymbol{a}}$ — обозначает нижнюю границу интервала, $\overline{\boldsymbol{a}}$ — верхнюю, $\underline{\boldsymbol{a}} \leq \overline{\boldsymbol{a}}$. Как $\mathbb{IR}$ будем обозначать заданное таким образом на $\mathbb{R}$ множество интервалов. Сумма интервалов $\boldsymbol{a}$ и $\boldsymbol{b}$ определяется следующим образом: $\boldsymbol{a} + \boldsymbol{b} = [\underline{\boldsymbol{a}} + \underline{\boldsymbol{b}}, \overline{\boldsymbol{a}} + \overline{\boldsymbol{b}}]$. Результат умножения интервала $\boldsymbol{a}$ на $\alpha \in \mathbb{R}_+ = \{a \in \mathbb{R} \mid a > 0\}$ — это интервал $\alpha\boldsymbol{a} = [\alpha\underline{\boldsymbol{a}}, \alpha\overline{\boldsymbol{a}}]$.

Интервал из $\mathbb{IR}$ является интервальным вектором размерности 1. Будем обозначать как $\mathbb{IR}^n$ множество n-мерных интервальных векторов с компонентами из $\mathbb{IR}$. Интервальный вектор $\boldsymbol{a}$ размерности n имеет своими компонентами интервалы $\boldsymbol{a}_i \in \mathbb{IR}$:

$$\boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = ([\underline{\boldsymbol{a}}_1, \overline{\boldsymbol{a}}_1], \ldots, [\underline{\boldsymbol{a}}_n, \overline{\boldsymbol{a}}_n]).$$

Далее нами рассматриваются интервалы из $\mathbb{IR}_+$, то есть такие $\boldsymbol{a} \in \mathbb{IR}$, что $\underline{\boldsymbol{a}} > 0$.

В задачах ИДО функция вида (1) задачи в постановке (I) заменяется на функцию вида

$$\boldsymbol{f}(x,\boldsymbol{c}) = \sum_{e\in E_x} \boldsymbol{c}(e) = \sum_{i=1}^{n} \boldsymbol{c}_i x_i, \tag{2}$$

где $\boldsymbol{c}_i = \boldsymbol{c}(e_i) \in \mathbb{IR}_+$ — интервалы стоимостей элементов E, $x = (x_1, \ldots, x_n) \in \mathcal{D}$.

Интервал $\boldsymbol{f}(x,\boldsymbol{c})$ — это интервал возможных значений стоимости решения $x \in \mathcal{D}$ для всех возможных реализаций из $\boldsymbol{c}$. Функция $\boldsymbol{f}$ задаётся с использованием определённых выше операций сложения интервалов и умножения их на число. Поскольку выражение целевой функции вида (1) содержит по одному вхождению каждой переменной в первой степени, то по основной теореме интервальной арифметики [2] для интервальной функции (2) имеем:

$$\boldsymbol{f}(x,\boldsymbol{c}) = \{f(x,c) \mid c \in \boldsymbol{c}\} = [f(x,\underline{\boldsymbol{c}}), f(x,\overline{\boldsymbol{c}})]. \tag{3}$$

3. Интервальный жадный алгоритм решения задачи о покрытии множества с интервальной целевой функцией

Достаточно общий подход к нахождению приближённых решений задач ДО в постановке (I) реализуется жадными алгоритмами. На итерациях жадного алгоритма строится множество Gr — набор элементов из E. Решение задачи формируется выбором элементов $e \in E$ в множество Gr с учётом их стоимостей $c(e)$ и других параметров индивидуальной задачи. Для этого используется функция выбора, с помощью которой на каждой итерации работы жадного алгоритма определяется элемент $e \in E$, добавление которого в Gr должно минимизировать стоимость получаемого решения. Работа жадного алгоритма завершается, когда на одной из его итераций получено такое множество Gr, что $Gr = E_x$ для некоторого $x \in \mathcal{D}$ — найденного с помощью жадного алгоритма решения задачи.

С помощью жадных алгоритмов может быть получено оптимальное решение задач ДО (I), множество допустимых решений которых $\mathcal{D}$ представляет собой матроид [3]. Для некоторых задач ДО жадные алгоритмы являются асимптотически лучшими полиномиальными алгоритмами решения ([4, 5] и др.).

В качестве примера задачи ДО в постановке (I) мы рассмотрим *задачу о покрытии множества* (далее — «ЗПМ»). В ЗПМ дано множество $\mathcal{U} = \{1, \ldots, m\}$. Дан набор его подмножеств $\mathcal{S} = \{S_1, \ldots, S_n\}$, $S_i \subseteq \mathcal{U}$, такой, что $\cup_{i=1}^{n} S_i = \mathcal{U}$. Набор множеств $\mathcal{S}' = \{S_{i_1}, \ldots, S_{i_k}\}$, $S_{i_j} \in \mathcal{S}$, называется *покрытием* $\mathcal{U}$, если $\cup_{j=1}^{k} S_{i_j} = \mathcal{U}$. Задан вектор $c = (c_1, \ldots, c_n) \in \mathbb{R}_+^n$ стоимостей $c_i = c(S_i)$ множеств $S_i \in \mathcal{S}$. Для набора множеств $\mathcal{S}' = \{S_{i_1}, \ldots, S_{i_k}\}$ его стоимость $c\,(S')$ равна сумме стоимостей входящих в него множеств: $c(\mathcal{S}') = \sum_{j=1}^{k} c(S_{i_j})$. Для решения ЗПМ необходимо найти *оптимальное покрытие*, то есть покрытие минимальной стоимости.

В постановке задачи (I) для ЗПМ набор подмножеств $\mathcal{S}$ множества $\mathcal{U}$ ассоциируется с множеством E, множество возможных покрытий $\mathcal{U}$ подмножествами из $\mathcal{S}$ — с множеством $\mathcal{D}$. Двоичный вектор $x \in \mathcal{D}$ размерности n задаёт покрытие $\mathcal{U}$: если $x_i = 1$, то S_i включается в покрытие, если $x_i = 0$ — не включается.

Обозначим как $c(x)$ стоимость покрытия, которое задаёт $x \in \mathcal{D}$. Заметим, что в неинтервальном случае имеем

$$c(x) = \sum_{\{i|x_i=1\}} c(S_i),$$

тогда как в интервальном случае, вообще говоря,

$$\boldsymbol{c}(x) \neq \sum_{\{i|x_i=1\}} \boldsymbol{c}(S_i),$$

где $\boldsymbol{c}(x)$ — это интервал возможных значений стоимости решения x для тех реализаций из $\boldsymbol{c}$, при которых x является оптимальным. При рассмотрении возможных приближённых решений ИЗПМ $\boldsymbol{c}(x)$ — это интервал возможных значений стоимости решения x для тех реализаций из $\boldsymbol{c}$, при которых x является α-приближённым решением ЗПМ, получаемым с помощью некоторого алгоритма. В обоих случаях $\boldsymbol{c}(x)$ включено в $\boldsymbol{f}(x, \boldsymbol{c})$, для которого верно (3), но, как правило, значение $\boldsymbol{c}(x)$ может быть уточнено, и тогда будет верным следующее включение:

$$\boldsymbol{c}(x) \subset \sum_{\{i|x_i=1\}} \boldsymbol{c}(S_i).$$

Пусть $\check{x}$ — оптимальное решение для ЗПМ. Обозначим как $x_{\text{А}}$ решение, получаемое с помощью алгоритма А. Положим $\rho(x_{\text{А}}) = c(x_{\text{А}})/c(\check{x})$. В [6] показано, что при условии $\text{P} \neq \text{NP}$ для любого полиномиального приближенного алгоритма А решения ЗПМ выполнено неравенство $\rho(x_{\text{А}}) > (1 - o(1)) \ln m$. Этот и другие результаты ([4, 7] и др.), касающиеся вычислительной сложности нахождения приближённых решений с заданной точностью для невзвешенной ЗПМ, то есть когда $c_i = 1$ для всех $i = \overline{1, n}$, распространяются и на случай взвешенной ЗПМ.

Жадный алгоритм для ЗПМ, имеющий вычислительную сложность $O(m^2 n)$, даёт решение $\tilde{x}$, для которого выполняется $\rho(\tilde{x}) = c(\tilde{x})/c(\check{x}) \leq H(m) \leq \ln m + 1$, где $H(m) = \sum_{k=1}^{m} 1/k$ [8]. Рассматриваемые нами далее приближённые решения ЗПМ — это приближённые решения, которые могут быть получены для неё с помощью жадного алгоритма.

В интервальной ЗПМ (далее — «ИЗПМ») заданы интервальные стоимости $\boldsymbol{c}_i = \boldsymbol{c}\,(S_i)$, то есть задано множество возможных реализаций $\boldsymbol{c} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n) \in \mathbb{IR}^n_+$. Оптимальным решением ИЗПМ будем считать *объединённое множество оптимальных решений* Σ — множество её возможных оптимальных решений:

$$\Sigma = \{x \in \mathcal{D} \mid (\exists c \in \boldsymbol{c})(f(x, c) = \min_{y \in \mathcal{D}} f(y, c))\}.$$

На сегодняшний день нет алгоритма, гарантированно находящего Σ, не производя полный поиск по всем реализациям из $\boldsymbol{c}$. Такой алгоритм может быть реализован только в отдельных случаях работы с дискретными интервалами. В [9] вводится понятие объединённого множества приближённых решений задачи ИДО. Множество, для некоторого $\alpha \in \mathbb{R}_+$ содержащее α-приближённые решения для всех возможных реализаций из $\boldsymbol{c}$, называется *объединённым множеством приближённых решений* задачи ИДО и обозначается как $\tilde{\Sigma}_\alpha$:

$$\tilde{\Sigma}_\alpha = \Big\{x \in \mathcal{D} \;\Big|\; (\exists c \in \boldsymbol{c}) \; \big(f(x, c) \leq \alpha \min_{y \in \mathcal{D}} f(y, c)\big)\Big\}.$$

То есть $\tilde{\Sigma}_\alpha$ — это такое множество решений из $\mathcal{D}$, что для любой реализации в нём содержится α-приближённое решение соответствующей этой реализации задачи ДО. Рассматривая далее ЗПМ с вещественными и интервальными коэффициентами, положим $\alpha = H(m)$ и будем обозначать объединённое множество приближённых решений как $\tilde{\Sigma}$.

ИЗПМ будут нами решаться нахождением множества $\tilde{\Sigma}$ с помощью интервального жадного алгоритма [9, 10]. Этот алгоритм представляет собой обобщение жадного алгоритма решения ЗПМ на случай интервальных стоимостей. В ходе его работы находятся подинтервалы заданных интервалов возможных значений стоимостей множеств, при которых они включаются в покрытие неинтервальным жадным алгоритмом. Используя эти оценки, мы получаем интервалы возможных значений стоимости $\boldsymbol{c}(\tilde{x})$ для $\tilde{x} \in \tilde{\Sigma}$. Также алгоритм даёт вероятности $\mathsf{P}(\tilde{x})$ получения решений $\tilde{x} \in \tilde{\Sigma}$ неинтервальным жадным алгоритмом для заданного на множестве реализаций $\boldsymbol{c}$ распределения.

Предложенный в [9, 10] подход может быть применён для получения интервальных жадных алгоритмов решения и других задач ДО в постановке (I). Вычислительная сложность интервального жадного алгоритма определяется мощностью множества $\tilde{\Sigma}$ и экспоненциальна в общем случае. В случае, если значение мощности множества $\tilde{\Sigma}$ ограничено полиномом от размерности задачи, вычислительная сложность интервального жадного алгоритма полиномиальна.

Рассмотрим индивидуальную ИЗПМ, для которой $m = 7$, $n = 11$. Набор $\mathcal{S}$ подмножеств $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7\}$ включает в себя следующие подмножества: $S_1 = \{3, 5\}$, $S_2 = \{4, 6\}$, $S_3 = \{1, 3\}$, $S_4 = \{2, 3, 4\}$, $S_5 = \{1, 5, 6\}$, $S_6 = \{4, 5, 6\}$, $S_7 = \{1, 4, 6, 7\}$, $S_8 = \{1, 3, 4, 6\}$, $S_9 = \{2, 4, 5, 7\}$, $S_{10} = \{1, 3, 6, 7\}$, $S_{11} = \{1, 2, 4, 6\}$. Средние значения стоимостей заданы вектором $c_\mu = (119, 117, 124, 135, 128, 130, 143, 144, 144, 142, 141)$. Предположим, что относительная погрешность оценок стоимостей c_i не превышает 5%, то есть $\boldsymbol{c}_i = [c_{\mu,i} - \delta, c_{\mu,i} + \delta]$, где $c_{\mu,i}$ — i-я компонента c_μ, $\delta = 5$.

Пусть $\tilde{x}^{(i)}$ обозначает одно из возможных приближённых решений. Множество возможных приближённых решений $\tilde{\Sigma}$ для рассматриваемой задачи приводится ниже. Также приводятся интервалы возможных значений стоимостей $\boldsymbol{c}(\tilde{x}^{(i)})$ и вероятности $\mathsf{P}(\tilde{x}^{(i)})$.

1) $\tilde{x}^{(1)} = (1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0)$, $\boldsymbol{c}(\tilde{x}^{(1)}) = [382, 410]$, $\mathsf{P}(\tilde{x}^{(1)}) = 0.1542$;
2) $\tilde{x}^{(2)} = (1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0)$, $\boldsymbol{c}(\tilde{x}^{(2)}) = [391, 410]$, $\mathsf{P}(\tilde{x}^{(2)}) = 0.0007$;
3) $\tilde{x}^{(3)} = (1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1)$, $\boldsymbol{c}(\tilde{x}^{(3)}) = [390, 410]$, $\mathsf{P}(\tilde{x}^{(3)}) = 0.1342$;
4) $\tilde{x}^{(4)} = (0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0)$, $\boldsymbol{c}(\tilde{x}^{(4)}) = [278, 295]$, $\mathsf{P}(\tilde{x}^{(4)}) = 0.1172$;
5) $\tilde{x}^{(5)} = (0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0)$, $\boldsymbol{c}(\tilde{x}^{(5)}) = [278, 293]$, $\mathsf{P}(\tilde{x}^{(5)}) = 0.3166$;
6) $\tilde{x}^{(6)} = (1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1)$, $\boldsymbol{c}(\tilde{x}^{(6)}) = [389, 417]$, $\mathsf{P}(\tilde{x}^{(6)}) = 0.0826$;
7) $\tilde{x}^{(7)} = (1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1)$, $\boldsymbol{c}(\tilde{x}^{(7)}) = [387, 417]$, $\mathsf{P}(\tilde{x}^{(7)}) = 0.1946$.

При решении задачи с помощью жадного алгоритма для реализации c_μ мы получим среднее приближённое решение $\tilde{x}_\mu = \tilde{x}^{(1)} = (1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0)$ со значением целевой функции 397. $\tilde{x}_\mu$ — лишь одно из семи возможных приближённых решений и имеет вероятность получения 0.1542, тогда как наиболее вероятным решением является $\tilde{x}^{(5)}$ с вероятностью получения 0.3166.

Заметим, что приближённое решение $\tilde{x}_\mu$ устойчиво, поскольку при $\delta \leq 0.5$ в δ-окрестности $c_\mu \in \mathbb{R}^n$ решение $\tilde{x}_\mu$ будет единственным возможным приближённым решением, получаемым с помощью жадного алгоритма. Однако с ростом радиуса интервальных ограничений на значения c_i количество возможных приближённых решений будет увеличиваться, как будут увеличиваться и различия в возможных значениях целевой функции для них. При $\delta = 5$ решение $\tilde{x}_\mu$ будет непредставительным уже для значения порога представительности $b=0.2$.

4. Индивидуальные задачи с непредставительным средним приближённым решением

4.1. Выборка из индивидуальных ИЗПМ, используемая при проведении эксперимента

Индивидуальные ИЗПМ, для которых в ходе вычислительного эксперимента мы находили объединённые множества приближённых решений, были сгенерированы нами по представленному ниже алгоритму. В ходе его выполнения нами генерируются множества из $\mathcal{S}$ — подмножества $\mathcal{U}$ случайной мощности со случайным набором элементов. Эти множества генерируются до тех пор, пока не будет получено такое $\mathcal{S}$, что каждый элемент $\mathcal{U}$ будет покрыт не менее чем q раз, то есть будет содержаться не менее чем в q множествах из $\mathcal{S}$. При проведении экспериментов, результаты которых приводятся далее, полагалось $q=3$.

Стоимости множеств из $\mathcal{S}$ задаются по следующей формуле:

$$c_i := 100 + 10 \cdot p_i + \eta, \tag{4}$$

где p_i — мощность S_i, η — равномерно распределённая на интервале $[-5, 5]$ целочисленная случайная величина. Пусть $S_i \in \mathcal{S}$ моделирует объект из некоторой предметной области, для которой ставится задача. По формуле (4) его стоимость c_i задаётся, главным образом, мощностью p_i. Случайная величина η моделирует возможные отличия в стоимостях объектов близких по значениям каких-то своих базовых характеристик.

Получив индивидуальную ЗПМ с коэффициентами $c=(c_1, \dots, c_n) \in \mathbb{R}^n_+$, мы моделируем неточность оценок (ошибки измерений) стоимостей c_i множеств $S_i \in \mathcal{S}$ интервалами $[c_i - \delta, c_i + \delta]$, то есть полагая $c_\mu = c$. При $\delta=5$ моделируемая относительная ошибка не превышает 5%.

Генерируя таким образом задачи, мы получаем выборку индивидуальных ИЗПМ. Генеральная совокупность в нашем эксперименте для фиксированных значений m и δ — это всевозможные пары, каждая из которых состоит:

1) из множества $\mathcal{P}^{(i)}(m, \delta)$, включающего в себя 1000 ИЗПМ, сгенерированных по представленному ниже алгоритму;
2) вектора $d^{(i)}(m, \delta)$, характеризующего распределение этих ИЗПМ в соответствии со значением вероятности $\mathsf{P}(\tilde{x}_\mu)$ получения среднего приближённого решения $\tilde{x}_\mu$:

$$d^{(i)}(m, \delta) = (d^{(i)}_1(m, \delta), \dots, d^{(i)}_{10}(m, \delta)),$$

где $d^{(i)}_k(m, \delta)$ — это доля (в процентах) задач из $\mathcal{P}^{(i)}(m, \delta)$, для которых

$$\mathsf{P}(\tilde{x}_\mu) \in ((k-1)/10, k/10].$$

Далее приводятся результаты усреднения этих значений для выборок

$$\{(\mathcal{P}^{(i)}(m, \delta), d^{(i)}(m, \delta))\}_{i=1}^{100}, \tag{5}$$

полученных для значений $m = 5, 10, 15, 20$, $\delta = \overline{1, 5}$. В табл. 1–8 представлены средние арифметические компонент $d^{(i)}(m, \delta)$, $i=\overline{1, 100}$, и их среднеквадратические отклонения.

Алгоритм генерирования индивидуальных ЗПМ.

Вход: m, q, δ.

1. $i := 0$.
2. Пока каждый элемент $\mathcal{U}$ покрыт менее q раз, выполнять:
   - 2.1. $i := i + 1$.
   - 2.2. Сгенерировать значение мощности p_i множества S_i: p_i — значение случайной величины, равномерно распределённой на множестве $\{1, \ldots, m\}$.
   - 2.3. Случайным образом в соответствии с равномерным распределением на $\mathcal{U}$ выбрать p_i элементов множества S_i из $\mathcal{U}$.
   - 2.4. Сгенерировать стоимость c_i множества S_i:
   $$c_i := 100 + 10 \cdot p_i + \eta,$$
   где η — равномерно распределённая на интервале $[-5, 5]$ целочисленная случайная величина.
   - 2.5. Сформировать интервальную стоимость $\boldsymbol{c}_i$ множества S_i: $\boldsymbol{c}_i := [c_i - \delta, c_i + \delta]$.

Выход: ИЗПМ с заданными $\mathcal{U}$, $\mathcal{S} = \{S_1, \ldots, S_n\}$, вектор $\boldsymbol{c} \in \mathbb{IR}^n_+$.

Мощность множества всех ИЗПМ, которые могут быть получены с помощью этого алгоритма, существенно превышает возрастающую экспоненциально с ростом m мощность множества всех наборов подмножеств, покрывающих q раз все m элементов $\mathcal{U}$ ($q \geq 2$) [11]. Хотя выборка из 100 элементов $\mathcal{P}^{(i)}(m, \delta)$ для фиксированных значений m и δ не может считаться большой, небольшая размерность решаемых задач позволяет получить статистически устойчивые результаты.

4.2. Вычислительные эксперименты

4.2.1. Доля индивидуальных ИЗПМ с непредставительным средним приближённым решением

При проведении экспериментов, результаты которых рассматриваются далее, для каждой сгенерированной нами ИЗПМ мы находили множество $\tilde{\Sigma}_\alpha$ с помощью интервального жадного алгоритма, $\alpha = H(m)$. По результатам этих вычислений сформированы приведённые ниже таблицы.

В строках табл. 1, 3, 5, 7 даны средние значения компонент векторов $d^{(i)}(m, \delta)$ из выборки (5) для $\delta = \overline{1, 5}$. Так, значение, приведённое в табл. 7 для $m = 20$, $\delta = 5$ и $k = 10$ означает, что, в среднем, доля задач из $\mathcal{P}^{(i)}(m, \delta)$, $i = \overline{1, 100}$, для которых $\mathsf{P}(\tilde{x}_\mu) \in (0.9, 1]$, составляет 36.63%.

Из этих результатов следует, что с ростом размерности задач m доля индивидуальных ИЗПМ с непредставительным средним приближённым решением растёт для всех значений порога представительности b. Уже при $m = 10$ среднее приближённое решение оказывается чаще непредставительным, нежели наоборот для $\delta = 5$, $b = 0.9$.

Т а б л и ц а 1. Средние значения компонент $d^{(i)}(5,\delta)$, $i=\overline{1,100}$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0	0.05	0.18	0.53	1.59	5.04	0.10	1.49	10.13	80.89
2	0	0.06	0.35	1.20	3.64	4.04	3.65	7.70	7.39	71.96
3	0	0.13	0.65	1.98	5.27	4.57	7.70	7.13	5.32	67.26
4	0	0.20	1.11	3.31	6.20	7.39	7.21	6.17	5.62	62.77
5	0	0.35	1.71	4.23	7.58	9.29	7.00	5.09	4.09	60.65

Т а б л и ц а 2. Среднеквадратические отклонения компонент $d^{(i)}(5,\delta)$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0	0.07	0.17	0.23	0.63	1.17	0.10	0.45	2.12	3.94
2	0.01	0.08	0.20	0.42	0.88	0.97	1.07	1.58	1.66	5.50
3	0.01	0.12	0.27	0.72	0.97	1.15	1.53	1.09	1.10	5.55
4	0.02	0.13	0.39	0.94	1.27	1.28	1.44	1.27	1.00	6.03
5	0.01	0.18	0.47	1.13	1.57	1.51	1.10	0.95	0.83	5.92

Т а б л и ц а 3. Средние значения компонент $d^{(i)}(10,\delta)$, $i=\overline{1,100}$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0	0.05	0.23	0.48	1.72	6.08	0.13	1.82	11.87	77.61
2	0	0.07	0.42	1.25	3.55	5.09	3.73	9.48	8.84	67.57
3	0.01	0.15	0.81	2.31	5.47	5.67	9.90	8.82	6.92	59.94
4	0.01	0.23	1.32	3.84	7.09	8.57	9.09	8.46	8.87	52.53
5	0.02	0.40	2.10	4.95	9.17	11.36	9.21	7.84	8.12	46.81

Т а б л и ц а 4. Среднеквадратические отклонения компонент $d^{(i)}(10,\delta)$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0.01	0.07	0.15	0.21	0.45	0.88	0.12	0.42	1.11	1.46
2	0	0.08	0.20	0.36	0.69	0.79	0.71	1.00	1.01	1.72
3	0.03	0.15	0.31	0.54	0.80	0.77	0.98	0.82	0.94	1.65
4	0.03	0.17	0.42	0.69	0.83	0.91	1.08	1.34	1.20	2.65
5	0.04	0.20	0.42	0.79	0.88	1.18	1.17	1.07	1.53	3.54

4.2.2. Рост размерности ИЗПМ и представительность возможных приближённых решений

Пусть P_μ обозначает среднее арифметическое вероятностей $\mathsf{P}(\tilde{x})$ по всем $\tilde{x} \in \tilde{\Sigma}$ для сгенерированной ИЗПМ, $\mathsf{P}_{\max}$ — максимальное значение $\mathsf{P}(\tilde{x})$ по всем $\tilde{x}\in\tilde{\Sigma}$. Обозначим

Т а б л и ц а 5. Средние значения компонент $d^{(i)}(15,\delta)$, $i=\overline{1,100}$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0	0.03	0.22	0.39	1.39	5.85	0.71	1.70	11.54	78.16
2	0	0.05	0.35	1.08	3.07	5.46	4.06	9.28	9.45	67.19
3	0	0.10	0.60	1.88	4.83	6.10	10.31	9.42	8.47	58.29
4	0.01	0.18	1.08	3.32	6.91	8.72	9.91	9.64	10.54	49.67
5	0.01	0.35	1.84	4.53	8.73	12.26	10.18	9.11	10.96	41.99

Т а б л и ц а 6. Среднеквадратические отклонения компонент $d^{(i)}(15,\delta)$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0	0.05	0.14	0.22	0.47	0.92	0.22	0.46	1.39	2.55
2	0.01	0.09	0.25	0.41	0.85	0.87	0.99	1.11	1.15	3.98
3	0.01	0.10	0.33	0.79	1.14	0.86	1.18	1.36	0.87	4.16
4	0.03	0.15	0.51	1.07	1.32	1.26	1.11	1.06	1.73	3.59
5	0.03	0.26	0.65	1.20	1.68	1.00	1.20	1.19	1.70	4.31

Т а б л и ц а 7. Средние значения компонент $d^{(i)}(20,\delta)$, $i=\overline{1,100}$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0	0.01	0.10	0.29	1.01	5.35	1.42	1.89	9.92	80.00
2	0	0.04	0.19	0.79	2.43	5.69	4.12	9.37	9.46	67.91
3	0	0.06	0.46	1.67	4.64	7.04	10.71	10.02	10.55	54.84
4	0	0.11	0.78	2.74	6.46	9.48	10.87	11.10	12.94	45.52
5	0.01	0.20	1.32	3.98	8.62	12.92	11.89	11.46	12.97	36.63

Т а б л и ц а 8. Среднеквадратические отклонения компонент $d^{(i)}(20,\delta)$.

$\delta\backslash k$	1	2	3	4	5	6	7	8	9	10
1	0.01	0.03	0.11	0.22	0.40	0.91	0.37	0.37	1.72	2.58
2	0	0.06	0.15	0.33	0.83	0.76	0.71	1.31	1.08	3.50
3	0	0.07	0.25	0.57	1.32	0.90	1.31	0.98	0.88	3.64
4	0.01	0.12	0.42	0.95	1.65	1.52	1.29	1.15	1.13	5.66
5	0.02	0.15	0.58	1.25	2.09	1.44	1.40	1.24	1.31	5.85

как MP_μ среднее арифметическое значений P_μ по всем сгенерированным ИЗПМ одной размерности m, как $\mathsf{MP}_{\max}$ — среднее арифметическое значений $\mathsf{P}_{\max}$ для них.

В табл. 9 представлены значения MP_μ и $\mathsf{MP}_{\max}$, рассчитанные по 1000 ИЗПМ, которые генерировались для каждого m. Для этих задач $\boldsymbol{c}_i = [c_i - \delta_i, c_i + \delta_i]$, где значение

$\delta_i = 0.05c_i$ определялось для каждого интервала стоимости индивидуально. Из представленных в таблице данных следует, что значения MP_μ убывают на один порядок при росте m на один порядок. В среднем для $m > 100$ даже значение $\mathsf{P}_{\max}$ оказывается меньше 0.25 для сгенерированных ИЗПМ, то есть среднее приближённое, как и любое другое возможное приближённое решение ИЗПМ, является непредставительным для $b \geq 0.25$.

Т а б л и ц а 9. Значения MP_μ и $\mathsf{MP}_{\max}$.

m	MP_μ	$\mathsf{MP}_{\max}$
100	0.0554	0.3675
200	0.0214	0.2636
300	0.0128	0.2235
400	0.0092	0.1984
500	0.0071	0.1794
600	0.0055	0.1656
700	0.0049	0.1603
800	0.0040	0.1498
900	0.0036	0.1420
1000	0.0033	0.1378

Заметим, что для значений $m \leq 20$, данные для которых представлены в табл. 1–8, доля задач с вероятностью получения среднего решения большей 0.9 достаточно велика и составляет, например, 36.63% при $m = 20$ и $\delta = 5$, тогда как при на порядок больших значениях m задач, где среднее приближённое или любое другое возможное приближённое решение было бы представительным при $b = 0.9$, как правило, нет.

4.2.3. Распределение стоимостей возможных приближённых решений

Вернёмся к индивидуальной ИЗПМ, объединённое множество приближённых решений $\tilde{\Sigma}$ которой мы рассматривали выше. Объединённое множество оптимальных решений для неё — это множество $\Sigma = \{\check{x}^{(1)}, \check{x}^{(2)}\}$, где $\check{x}^{(1)} = (0,0,0,0,0,0,0,1,1,0,0)$ и $\check{x}^{(2)} = (0,0,0,0,0,0,0,0,1,1,0)$. $f(\check{x}^{(i)}, c) \in [276, 296]$ для всех $c \in \boldsymbol{c}$, $i = \overline{1,2}$. Имеем $\mathsf{P}(\check{x}_\mu) = 0.667$, $f(\check{x}_\mu, c_\mu) = 286$. То есть $\check{x}_\mu$ является наиболее вероятным и максимальное отклонение стоимости возможного приближённого решения от $f(\check{x}_\mu, c_\mu)$, то есть от стоимости оптимального решения для средних значений коэффициентов, по всем реализациям составит 10:

$$\max_{\substack{c \in \boldsymbol{c} \\ \check{x} \in \Sigma}} |f(\check{x}_\mu, c_\mu) - f(\check{x}, c)| = 10. \tag{6}$$

На рис. 1 представлена гистограмма частот (ось p) получаемых значений стоимости (ось f) возможных приближённых решений той же ИЗПМ по результатам решения индивидуальных ЗПМ с неинтервальной целевой функцией для 10^6 реализаций $c \in \boldsymbol{c}$, сгенерированных в соответствии с равномерным распределением на $\boldsymbol{c}$. Гистограмма приближает график функции плотности распределения стоимости возможных

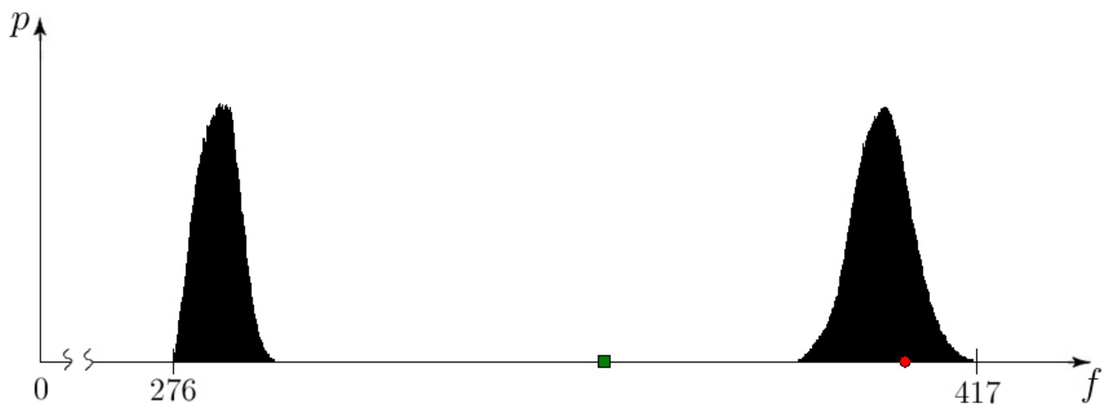

Рис. 1. Гистограмма частот стоимостей возможных приближённых решений

приближённых решений при заданном равномерном распределении на множестве реализаций $\boldsymbol{c}$. На оси f отмечена средняя стоимость возможного приближённого решения, рассчитанная как математическое ожидание по следующей формуле и примерно равная 350.51:

$$\sum_{\tilde{x}\in\tilde{\Sigma}} \mathsf{P}(\tilde{x}) \cdot \frac{\underline{\boldsymbol{c}}(\tilde{x}) + \overline{\boldsymbol{c}}(\tilde{x})}{2} \approx 350.51.$$

Это значение отмечно квадратом на оси f. Также на оси f отмечена стоимость среднего приближённого решения $\tilde{x}_\mu$ для средних значений коэффициентов: значение $f(\tilde{x}_\mu, c_\mu)$ равное 397 отмечено кругом на оси f. Заметим, что для рассматриваемой нами индивидуальной ИЗПМ множество $\{f(\tilde{x}, c) \mid (\tilde{x} \in \tilde{\Sigma})(c \in \boldsymbol{c})\}$ возможных значений целевой функции для решений из $\tilde{\Sigma}$ несвязно.

Можно видеть, что значение $f(\tilde{x}_\mu, c_\mu) = 397$ сильно отклонено от средней по всем реализациям из $\boldsymbol{c}$ стоимости возможного приближённого решения, и

$$\max_{\substack{c\in\boldsymbol{c}\\ \tilde{x}\in\tilde{\Sigma}}} |f(\tilde{x}_\mu, c_\mu) - f(\tilde{x}, c)| = 121. \tag{7}$$

Сравнивая (6) и (7), отметим, что стоимости возможных приближённых решений ИЗПМ могут отличаться от стоимости приближённого решения для средних значений коэффициентов, гораздо больше, чем стоимости возможных оптимальных решений той же задачи отличаются от стоимости оптимального решения для средних значений коэффициентов.

Этот пример показывает, что хотя среднее оптимальное решение ИЗПМ может быть представительным при некоторых значениях порога представительности b, среднее приближённое, как и любое другое возможное приближённое решение задачи, могут при этом оказываться непредставительными для значительно меньших значений b. Так, в рассматриваемой ИЗПМ среднее приближённое решение $\tilde{x}_\mu$ оказывается непредставительным уже при $b = 0.155$. Значение же целевой функции $f(\tilde{x}_\mu, c_\mu)$ для него оказывается ближе к максимально возможному значению стоимости приближённого решения, нежели к его среднему значению.

5. Выводы

Приведённые в работе результаты вычислительных экспериментов показывают, что средние приближённые решения, как и любые другие возможные приближённые решения задач дискретной оптимизации, могут быть непредставительными. Из этого следует, что при использовании методов дискретной оптимизации для решения прикладных задач, параметры которых заданы неточно, требуется обоснование представительности получаемых с их помощью возможных (оптимальных или приближённых) решений. Если такое обоснование дать нельзя, то необходимо решать задачу с учётом имеющихся неопределённостей в её постановке. Это можно делать, например, получая и анализируя объединённое множество возможных оптимальных решений или, в случае невозможности его получения, получая и анализируя объединённое множество приближённых решений задачи и соответствующих им множеств возможных значений целевой функции.

Автор благодарит В. А. Мотовилова — выпускника математического факультета Омского государственного университета, чья помощь позволила сократить более чем в 5 раз время вычислений, необходимых для написания этой работы. Общее время, потребовавшееся для получения представленных в работе результатов, составило около 700 часов на ПК.

Список литературы / References

[1] **Гэри М., Джонсон Д.** Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. 419 с.

[2] **Шарый С. П.** Конечномерный интервальный анализ. Адрес доступа: http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf (дата обращения 1.03.2022).

[3] **Papadimitriu C. H., Steiglitz K.** Cobinatorial optimization: Algorithms and Complexity, Prentice Hall. 1987. 496 p.

[4] **Feige U.** A threshhold of $\ln n$ for approximating set cover // Journal of the ACM. 1998. V. 45 (4). P. 634–652.

[5] **Frieze A., Szpankowski W.** A greedy algorithm for the shortest common superstring is asymptotically optimal. Algorithmica. 1998; 21 (1). P. 21–36.

[6] **Dinur I., Steurer D.** Analytical approach to parallel repetition. // STOC'14: Proceedings of the forty-sixth annual ACM symposium on Theory of computing, ACM. 2013. P. 624–633.

[7] **Raz R., Safra S.** A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP // STOC'97: Proceedings of the twenty-ninth annual ACM symposium on Theory of computing. 1997. P. 475–484.

[8] **Chvatal V.** A greedy heuristic for the set-covering problem // Mathematics of operation research. 1979. V. 4 (3). P. 233–235.

[9] **Пролубников А. В.** Задача о покрытии множества с интервальными весами подмножеств и жадный алгоритм её решения // Вычислительные технологии. 2015. Т. 20 (6). С. 70–84.

[10] **Пролубников А. В.** Об одном подходе к решению задачи о покрытии с интервальными весами и его вычислительной сложности // Вычислительные технологии. 2017. Т. 22 (2). С. 115–126.

[11] **Devitt J. S., Jackson D. M.** The enumerations of covers of a finite set // Journal of the London Mathematical Society. 1982. V. (2) 25. P. 1–6.