

## ИССЛЕДОВАНИЕ АЛГОРИТМОВ РЕШЕНИЯ ОБОБЩЕННОЙ ЗАДАЧИ О МИНИМАЛЬНОМ ПОКРЫТИИ

© 2014 Б.А.Есипов, В.В.Муравьев

Самарский государственный аэрокосмический университет имени академика С.П. Королёва  
(национальный исследовательский университет)

Поступила в редакцию 17.12.2013

В данной статье рассмотрена математическая модель обобщенной задачи о минимальном покрытии. Предложены методы её точного и приближенного решения, исследованы сложность, быстродействие и точность.

Ключевые слова: математическая модель, минимальное покрытие множества, аддитивный алгоритм, генетический алгоритм.

Научный анализ организационно-технических систем приводит к необходимости математического моделирования и решения сложных задач оптимизации. В частности, при постановке задач наиболее эффективного распределения ресурсов по отдельным объектам, а также в задачах управления организационно-техническими системами возникает так называемая задача об оптимальном покрытии. Например, мы хотим обеспечить всю совокупность функций при производстве на отдельных объектах: цехах, участках за счет установки оборудования с частично перекрывающимися функциями. Нужно, чтобы все заданные функции на каждом объекте были обеспечены и при этом затраты на оборудование были бы минимальны. Аналогичные задачи возникают при кадровом обеспечении удаленных воинских частей, комплектовании космических аппаратов, кораблей, больниц, выбора инструментов для ремонтов или при проектировании систем на основе стандартных наборов микропроцессоров и электронных схем. Известны практические постановки по оптимизации постов ГАИ, состава меню программных интерфейсов, минимизации булевых функций и т.п. Для понимания существа подобных задач приведем следующую постановку.

Имеется  $m$  туристических групп, в каждой из которых известен набор языков общения и  $n$  переводчиков, каждый из которых может обеспечить определенный набор языков общения из общего числа  $l$  языков. Необходимо распределить переводчиков по туристическим группам так, чтобы обеспечить общение во всех группах и при этом чтобы количество занятых переводчиков

*Есипов Борис Алексеевич, кандидат технических наук, доцент кафедры информационных систем и технологий. E-mail: bobpereira@yandex.ru*  
*Муравьев Вячеслав Вячеславович, студент факультета информатики.*

было минимально. Можно ввести стоимости переводчиков  $c_{ij} \geq 0$  и минимизировать общие затраты на переводчиков. Такую задачу назовем *обобщенной* задачей о минимальном покрытии (ОЗМП). Легко заметить, что если в этой задаче только одна туристическая группа ( $m=1$ ), то получим обычную задачу о минимальном покрытии (ЗМП). В [2] содержится обзор методов решения ЗМП и рассмотрены вопросы сложности ее решения. В [1] предложена математическая модель ОЗМП в виде задачи линейного булева программирования, решение которой приводит к сложным переборным алгоритмам.

В данной работе предложен подход к решению ОЗМП, запрограммировано несколько алгоритмов ее решения и сделан анализ их быстродействия для задач большой размерности.

Математическая постановка ОЗМП отличается от ЗМП наличием нескольких независимых покрываемых множеств и заключается в следующем.

Пусть дано множество  $K = (1, 2, 3, \dots, k, \dots, l)$  и два набора его подмножеств  $A_1, \dots, A_n$  и

$B_1, \dots, B_m$ , таких, что  $\bigcup_{i=1}^m B_i \subseteq \bigcup_{j=1}^n A_j$ . Совокупность подмножеств  $A_j, j \in J_i \subseteq \{1, \dots, n\}$ , называется покрытием множества  $B_i$ , если  $\bigcup_{j \in J_i} A_j = B_i$ ;  $J_i$  в

таком случае называется множеством покрывающих индексов. Каждой паре подмножеств  $(B_i, A_j)$  приписан вес  $c_{ij} \geq 0$ . Весом покрытия

$A_j$  множества  $B_i$  является  $\sum_{j \in J_i} c_{ij}$ . Требуется найти

покрытие для всех множеств из набора  $B_1, \dots, B_m$  минимального суммарного веса такое, что множества покрывающих индексов данных покрытий  $J_i$  попарно не пересекаются. В терминах булева программирования эта задача формулируется следующим образом: найти

$$f^*(c, A, B) = \min \left\{ \sum_{j=1}^n \sum_{i=1}^m c_{ji} x_{ji} \mid Ax \geq B, x_{ji} \in \{0,1\} \right\},$$

при условии

$$\sum_{i=1}^m x_{ji} \leq 1, \quad j = 1, n. \quad (1)$$

Здесь  $A = (a_{kj})$  – матрица размера  $l \times n$  с элементами  $a_{kj} = 1$ , если  $k \in A_j$ , и  $a_{kj} = 0$  в противном случае;  $B = (b_{ki})$  – матрица размера  $l \times m$  с элементами  $b_{ki} = 1$ , если  $k \in B_i$ , и  $b_{ki} = 0$  в противном случае;  $c = (c_{ji})$  – матрица весов;  $x = (x_{ji})$  – матрица переменных с компонентами  $x_{ji} = 1$ , если  $A_j$  входит в покрытие  $B_i$ , и  $x_{ji} = 0$  в противном случае.

В отличие от классической задачи о минимальном покрытии (ЗМП), где количество допустимых векторов с элементами  $x_j$  равно  $2^n$ , то есть любой вектор является допустимым, в обобщенной задаче (ОЗМП) количество матриц с элементами  $x_{ij}$ , (количество возможных состояний), много меньше чем  $2^{mn}$  и равно  $(m+1)^n$ .

В самом деле, ограничение (1) означает, что в каждом столбце матрицы  $x$  может быть не более одной единицы, то есть каждый столбец представим одним из  $m+1$  способов: все элементы столбца равны 0 – 1 способ, или один из элементов равен 1 –  $m$  способов. В матрице  $x$   $n$  столбцов, значения в каждом столбце не зависят от значений в других столбцах, откуда матрица представима  $(m+1)^n$  способами. Для любого целого  $m$   $2^{mn} \geq (m+1)^n$ , причем равенство достигается только при  $m = 1$ .

В данной работе предлагается переход от булевых к целочисленным переменным:

$$f^*(c, A, B) = \min \left\{ \sum_{j=1}^n c(x_j, j) \mid Ah \geq B, x \in \{0, \dots, m\}^n \right\}$$

Здесь  $A = (a_{kj})$  – матрица размера  $l \times n$  с элементами  $a_{kj} = 1$ , если  $k \in A_j$ , и  $a_{kj} = 0$  в противном случае;  $B = (b_{ki})$  – матрица размера  $l \times m$  с элементами  $b_{ki} = 1$ , если  $k \in B_i$ , и  $b_{ki} = 0$

в противном случае;  $c(j, i) = \begin{cases} c_{ji}, & i \neq 0 \\ 0, & i = 0 \end{cases}$  – весовая функция,  $(c_{ji})$  – матрица весов;  $x = (x_j)$  – вектор целочисленных переменных с компонентами  $x_j = i$ , если  $A_j$  входит в покрытие  $B_i$ , и  $x_j = 0$ , если  $A_j$  не входит ни в одно покрытие;  $h = (h_{ji})$  – дополнительная матрица,  $h_{ji} = 1$ , если  $x_j = i$ , и  $h_{ji} = 0$  в противном случае.

Задача в такой формулировке не имеет недопустимых состояний, так как ограничение (1) выполняется в данном случае автоматически. Такая модификация позволяет не делать лишних проверок и переходов в процессе решения, что положительно сказывается на времени поиска.

**Замечание 1.** При проверке ограничения  $Ah \geq B$  в матрице произведения  $D = Ah$  имеет смысл вычислять элементы  $d_{ki}$  только с такими индексами  $(k, i)$ , что  $b_{ki} = 1$ . Это означает, что в задаче присутствует не  $l \times m$  ограничений, а

$$\text{только } p = \sum_{k=1}^l \sum_{i=1}^m b_{ki}, [1].$$

**Замечание 2.** Хотя задача проверки существования решения для данной задачи аналогична самой задаче минимального покрытия, существует ограничения на входные данные, невыполнение которых обозначает невозможность решения данной задачи. В данной работе применяется следующее ограничение:

$$\sum_{k=1}^l \sum_{j=1}^n a_{kj} \geq \sum_{k=1}^l \sum_{i=1}^m b_{ki}$$

– суммарное кол-во элементов множества  $K$ , входящих в набор множеств  $A_1, \dots, A_n$  больше суммарного кол-ва элементов множества  $K$ , входящих в набор множеств  $B_1, \dots, B_m$ , [1].

Рассмотренная задача относится к классу NP-полных задач и в общем случае для нее существует лишь экспоненциальное переборное решение.

Для разработки алгоритма решения ОЗМП можно применить аддитивный метод, генетический алгоритм, алгоритм муравьиной колонии, методы “лагранжевой релаксации” и другие [2]. Для разработки алгоритмов решения ОЗМП мы выбрали аддитивный и генетический методы дискретной оптимизации. Первый дает точное решение и использует зондирование решений с оценкой подмножества частичных решений: [3]. Генетический алгоритм использует методы естественной эволюции, такие как наследование, мутации, отбор и кроссинговер. Отличительной особенностью генетического алгоритма является акцент на использование оператора “скрещивания”, который производит операцию рекомбинации решений-кандидатов, роль которой аналогична роли скрещивания в живой природе [4]. При реализации этих алгоритмов использованы вышеописанные модификации выбора решений.

Для исследования аддитивного и генетического алгоритмов была написана программа на языке Java 7. В ней генерировались варианты исходных данных с заданными свойствами, каждый вариант задачи решался на отдельном алгоритме, затем результаты обобщались. Входными данными для программы являются величины:  $m, n, l$ , матрицы  $a_{kj}, b_{ki}$  и  $c_{ji}$ . Выходными данными программы является итоговая стоимость покрытия  $\sum_{j=1}^n c(x_j, j)$  и множества по-

крывающих индексов  $J_i, i = 1, m$

Ниже приведены результаты тестирования, выполненные на компьютере Asus, Core i5-3210M 2.5 GHz, 4 GB RAM.

Для всех тестов величина  $l$  была постоянной и равнялась 10; матрицы весов заполнялись случайными числами в промежутке  $[0;1000]$ , матрицы  $a_{kj}, b_{ki}$  заполнялись также случайно.

Для данного тестирования было выбрано несколько промежутков количества возможных состояний решения. Каждый порядок был в 10 раз больше предыдущего. Для каждого промежутка было сгенерировано множество входных данных, на которых запускались оба алгоритма, после чего сравнивалось время их работы. Для генетического алгоритма дополнительно подсчитывалось относительное отклонение ответа от точного, полученного аддитивным алгоритмом, а также процент неудачных попыток найти решение.

Как и следовало ожидать, аддитивный алгоритм находит ответ быстрее, чем генетический на тех же данных. С другой стороны, аддитивный

алгоритм более затратен в плане используемой памяти, что не позволяет использовать его при решении задач с большим количеством состояний. Недостатком генетического алгоритма является его рандомизированность. Из рис.2 и 3 видно, что с ростом количества состояний решения увеличиваются относительное отклонение результата от минимального и количество ложных “отсутствий решения” задачи.

Для сравнения времени работы генетического алгоритма на различных входных данных было выбрано несколько промежутков количества возможных состояний решения, каждый порядок был в 10 раз больше предыдущего и для каждого промежутка было сгенерировано множество входных данных, на которых запускался алгоритм. Здесь количество возможных состояний было на порядки больше, чем при тестировании аддитивного алгоритма.

Из рис. 4. видно, что, в среднем, при росте порядка количества возможных состояний решения время работы генетического алгоритма дос-

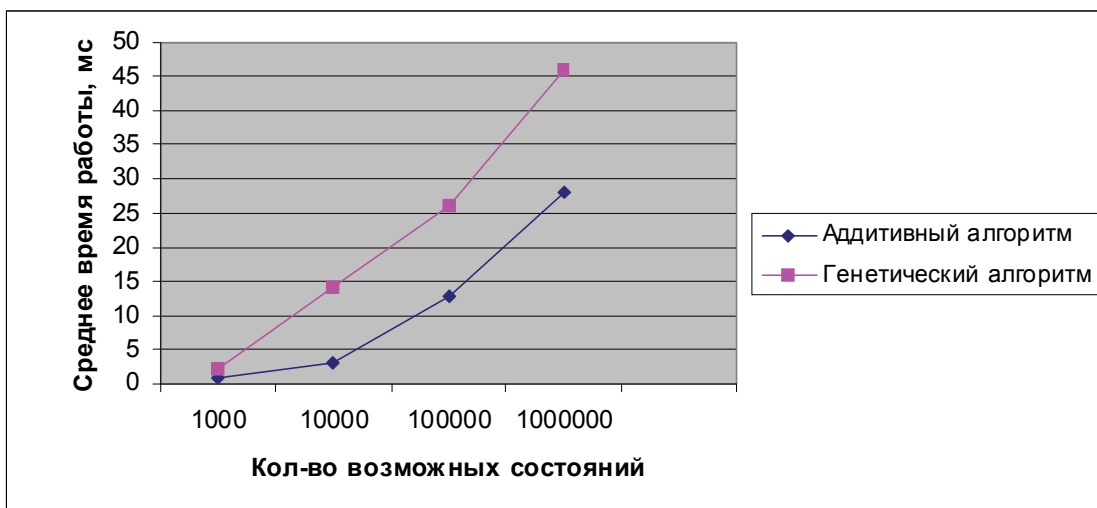


Рис. 1. Зависимость среднего времени работы алгоритмов от количества возможных состояний решения

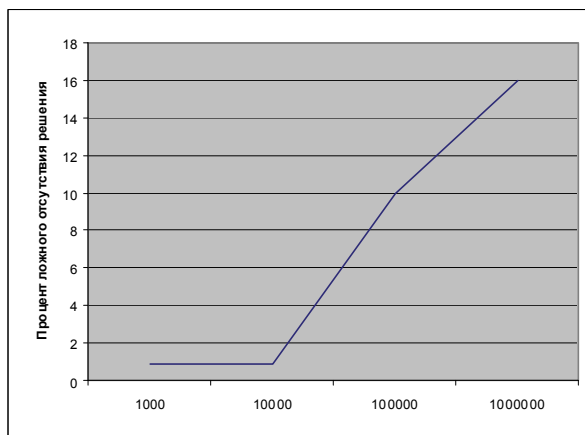
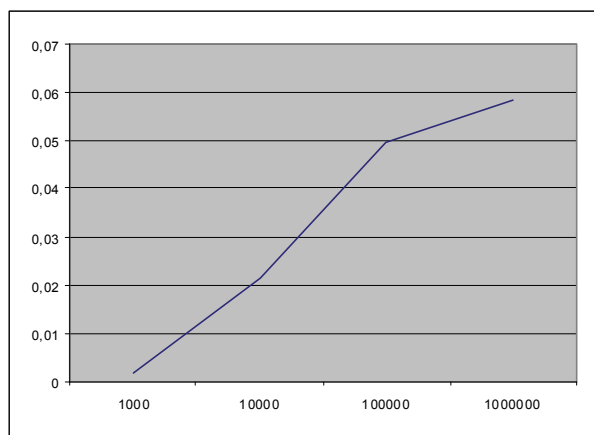


Рис. 2-3. Зависимость среднего относительного отклонения ответа и количества ложного “отсутствия решения” генетического алгоритма от количества возможных состояний решения

тигает определенного диапазона, в пределах которого изменяется. Это можно объяснить тем, что при большом порядке количества состояний алгоритм или быстрее начинает сходиться к правильному ответу при  $n \gg t$ , или быстрее определяет отсутствие решения при  $n \approx t$ .

Для анализа возможностей генетического алгоритма были выбраны несколько различных размеров популяции, начиная с 10 и заканчивая 1000. Во всех тестах количество возможных состояний находилось в промежутке  $[10^5; 10^6]$ , для каждого теста искалось контрольное решение с помощью аддитивного алгоритма и решения для каждого размера популяции.

На рис.5–7 показаны графики зависимости различных характеристик работы генетического алгоритма от размера популяции. Видно, что размер популяции порядка 100-250 является оп-

тимальным – меньшие размеры дают слишком неточные результаты, большие – требуют намного больше времени для нахождения результата.

### ЗАКЛЮЧЕНИЕ

Предложенная модификация математической модели и алгоритмов решения обобщенной задачи о минимальном покрытии дает существенную экономию в размерности решаемой задачи как аддитивным методом, дающим точное решение, так и генетическим алгоритмом, гарантирующим лишь приближенное решение. Разработанный комплекс программ тестирования и исследование быстродействия и точности решения задачи показывают практическую эффективность применения модели ОЗМП и компьютерных программ.



Рис. 4. Среднее время работы генетического алгоритма

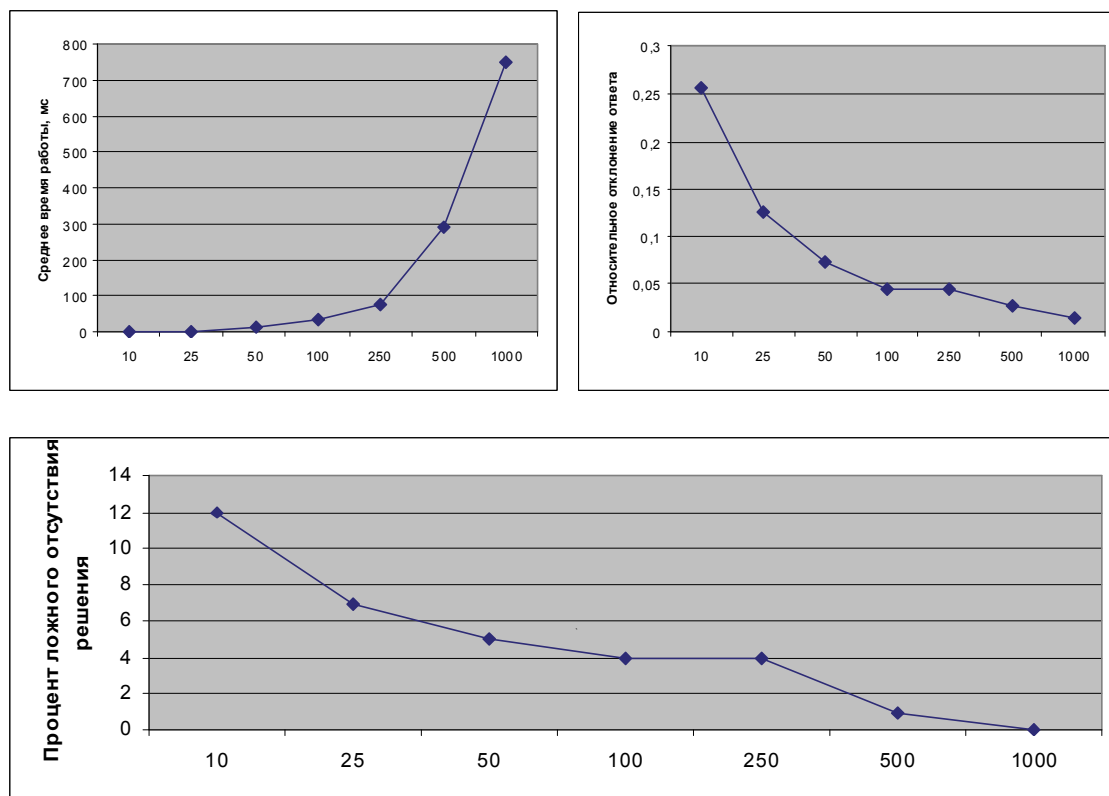


Рис. 5-7. Характеристики генетического алгоритма

### СПИСОК ЛИТЕРАТУРЫ

1. *Есипов Б.А.* Математическая модель и решение обобщенной задачи о покрытии // Избранные труды Международной конференции с элементами научной школы для молодежи “Перспективные информационные технологии для авиации и космоса” (ПИТ-2010). Самара, 2010. С.70-71.
2. Задача о покрытии множества: сложность, алгоритмы, экспериментальные исследования / *А.В. Еремеев, Л.А. Заозерская, А.А. Колоколов* // Дискретный анализ и исследование операций. Июль–декабрь 2000. Серия 2. Т. 7. № 2. С.22-46.
3. *Ху Т.* Целочисленное программирование и потоки в сетях. М.: Мир, 1974. 520 с.
4. *Нгуен Минь Ханг.* Применение генетического алгоритма для задачи нахождения покрытия множества // Труды института системного анализа РАН. 2008. Т. 33. С. 206-219.

### STUDY OF ALGORITHMS FOR SOLVING THE GENERALIZED PROBLEM OF THE MINIMAL COVERING

© 2014 B.A.Yesipov, V.V. Muraviev

Samara State Aerospace University named after Academician S.P. Korolyov  
(National Research University)

In this article we described mathematic model of the generalized problem of the minimal covering. Given methods for solving this problem and made an analysis of its degree of sophistication, performance and presigion.

Key words: mathematic model, minimal covering , additive algorithm, genetic algorithm.