ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь—июнь 2004. Серия 2. Том 11, № 1, 3–10

УДК 517.95

АЛГОРИТМЫ С УЛУЧШЕННЫМИ ОЦЕНКАМИ ТОЧНОСТИ ДЛЯ ЗАДАЧИ О ПОКРЫТИИ МНОЖЕСТВАМИ*)

А. А. Агеев

Рассматривается классическая задача о покрытии множествами: для заданного конечного множества I и семейства его подмножеств $\{S_j \mid j \in J\}$ с приписанными неотрицательными весами w_j требуется найти подсемейство $\{S_j \mid j \in J^*\}$ с минимальным суммарным весом среди всех подсемейств, объединение которых совпадает с I. В работе предлагаются алгоритмы с улучшенными оценками точности для некоторых NP-трудных частных случаев этой задачи.

Введение

В задаче о покрытии множествами (далее Set Cover) заданы конечное множество I и семейство его подмножеств $\{S_j|j \in J\}$ такое, что $\bigcup\limits_{j\in J} S_j = I$. Каждому множеству S_j приписан неотрицательный вес w_j. Требуется найти подсемейство $\{S_j|j \in J^*\}$ с минимальным суммарным весом среди всех подсемейств, объединение которых совпадает с I (иначе говоря, покрывает все элементы множества I). В дальнейшем будем считать, что $I = \{1, \dots, m\}$, $J = \{1, \dots, n\}$. Задача Set Cover легко может быть переформулирована как задача целочисленного линейного программирования. Для этого введем булеву матрицу $A = (a_{ij})$ размера $m \times n$, определенную следующим образом: $a_{ij} = 1$, если $i \in S_j$, и $a_{ij} = 0$ в противном случае. Каждому подмножеству S_j поставим в соответствие булеву переменную x_j по правилу: $x_j = 1$, если S_j входит в покрытие, и $x_j = 0$ в противном случае. В введенных обозначениях задача Set Cover эквивалентна следующей задаче: найти

$$\min \quad F(x) = \sum_{j\in J} w_j x_j \tag{1}$$

*) Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проекты 01–01–00786 и 02–01–01153) и INTAS (проект 00–217).

при ограничениях:

$$\sum_{j\in J} a_{ij}x_j \geqslant 1, \qquad i \in I, \tag{2}$$

$$x_j \geqslant 0, \qquad j \in J, \tag{3}$$

$$x_j \in \{0,1\}, \quad j \in J. \tag{4}$$

Для формулировки других результатов потребуются обозначения для некоторых параметров матрицы A:

- $r(A)$ — максимальное число единиц в строке матрицы A;
- $c(A)$ — максимальное число единиц в столбце матрицы A;
- $d(A)$ — максимальное число блоков из последовательных единиц в строке матрицы A.

Известно, что задача Set Cover является NP-трудной [7]. Более того, установлено [6], что из существования полиномиального алгоритма для ее решения с оценкой точности $(1-o(1))\ln m$ вытекает, что задачи из класса NP могут быть решены алгоритмами с трудоемкостью $n^{O(\log\log n)}$ (что принято считать столь же невероятным, как NP=P). С другой стороны, известно [5], что жадный алгоритм решает задачу Set Cover с оценкой точности

$$\sum_{t=1}^{c(A)} \frac{1}{t} \leqslant \ln m + 1,$$

т. е. для Set Cover жадный алгоритм в определенном смысле имеет наилучшую из возможных оценок точности (более точная оценка в терминах m установлена в [14]).

Если $c(A) = 2$, то задача полиномиально разрешима; она сводится к задаче о покрытии вершин графа ребрами. Для каждого фиксированного значения $c(A) \geqslant 3$ задача NP-трудна [7], но при этом, как нетрудно видеть, упомянутый выше жадный алгоритм применительно к этому частному случаю имеет константную оценку точности. С другой стороны, Set Cover может быть решена малотрудоемким двойственным алгоритмом с оценкой точности $r(A)$ [4]. Для каждого значения $r(A)$ эта оценка, по-видимому, неулучшаема. По крайней мере, для задачи о вершинном покрытии, соответствующей случаю $r(A) = 2$, до настоящего времени не доказана, но и не опровергнута гипотеза [11], утверждающая, что существование алгоритма с оценкой точности $2 - \varepsilon$ влечет NP=P.

В [8] рассмотрена задача о протыкании прямоугольников минимальным числом прямых, параллельных осям. Эта задача легко сводится к Set Cover с матрицей $A = (A'|A'')$, где A' и A'' — матрицы с $d(A) = 1$.

Известно [9], что эта задача NP-трудна. В [8] построен алгоритм с оценкой точности 2 для решения данной задачи, основанный на оригинальной идее округления линейной релаксации.

Основными результатами настоящей работы являются два новых приложения расширенного варианта этого метода. Первое приложение — алгоритм General с оценкой точности $d(A)$ для задачи о покрытии множествами в общей постановке. Ясно, что $d(A)$ всегда не превосходит $r(A)$, но в отличие от $r(A)$ параметр $d(A)$ зависит от перестановки столбцов матрицы A. В конкретных примерах и некоторых частных случаях $d(A)$ (например, когда $d(A) \leqslant$ const) может иметь существенно меньшее значение. Этот алгоритм обобщает алгоритм, построенный в [8].

Второе приложение имеет результатом алгоритм Special, который также обобщает алгоритм, построенный в [8], но обеспечивает лучшую оценку точности (почти двукратное улучшение, если сравнивать с предыдущим алгоритмом) для одного интересного частного случая Set Cover.

1. Базовая лемма и общая схема метода

Внимательный анализ предложенного в [8] алгоритма приводит к заключению, что его ключевая идея содержится в следующей лемме.

Рассмотрим индивидуальную задачу Set Cover с матрицей ограничений A.

Лемма 1. *Пусть x — оптимальное решение линейной релаксации и $A' = (a'_{ij})(i \in I, j \in J)$ — булева матрица, удовлетворяющая условиям:*

1. *Множество единиц матрицы A' является подмножеством множества единиц матрицы A, т. е.*

$$a'_{ij} \leqslant a_{ij} \tag{5}$$

для любых $i \in I$ и $j \in J$.

2. *Существует число $\alpha(A) > 0$ такое, что при любом $i \in I$*

$$\sum_{j \in J} a'_{ij} x_j \geqslant \frac{1}{\alpha(A)}. \tag{6}$$

3. *Линейная релаксация задачи* Set Cover *с матрицей ограничений A' имеет целочисленное оптимальное решение x'.*

Тогда x' — допустимое решение исходной задачи и

$$F(x') \leqslant \alpha(A) F(x^*)$$

для любого допустимого решения x^* исходной задачи.

Доказательство. Из условия 2 вытекает, что в каждой строке матрицы A' имеется по крайней мере одна единица. Это обстоятельство вместе с условием 1 влечет допустимость решения x' для исходной задачи. С другой стороны, из (6) следует, что $\alpha(A)x$ — допустимое решение задачи с матрицей A'. Следовательно, $F(x') \leqslant F(\alpha(A)x)$. В силу линейности F имеем $F(\alpha(A)x) \leqslant \alpha(A)F(x)$, что по определению x не превосходит $\alpha(A)F(x^*)$ для любого допустимого решения x^* исходной задачи. Лемма 1 доказана.

Опишем общую схему эффективного применения этой леммы. Пусть частный случай задачи Set Cover определяется некоторым классом матриц ограничений $\mathcal{A}$. Предположим, что по любой матрице $A \in \mathcal{A}$ и любому оптимальному решению линейной релаксации задачи Set Cover с матрицей ограничений A за полиномиальное время можно построить полностью унимодулярную матрицу A', удовлетворяющую условиям 1 и 2. Тогда любое оптимальное решение x' задачи Set Cover с матрицей A' (заметим, что оно целочисленно) будет одновременно оптимальным решением линейной релаксации этой задачи, т. е. условие 3 также выполняется. Поскольку задача Set Cover с полностью унимодулярной матрицей сводится к линейной релаксации [10, 12], x' можно найти за полиномиальное время. Таким образом, для данного частного случая Set Cover мы получаем полиномиальный приближенный алгоритм, имеющий согласно лемме 1 оценку точности $\alpha(A)$.

В следующих разделах мы представим новые реализации этой схемы.

2. Алгоритм для общего случая

Заметим, что при $d(A) = 1$ задача полиномиально разрешима (за время $O(mn)$, если свести задачу к задаче минимизации правильного полинома от булевых переменных (см. [2]) и затем применить алгоритм из [3]). Более того, в этом случае матрица A является полностью унимодулярной [13, с. 540–544].

Пусть $U_i = \{j \in J \mid a_{ij} = 1\}$, т. е. U_i есть множество столбцов, имеющих единицу в пересечении с i-й строкой. Множество U_i можно представить как объединение $U_{1i} \cup U_{2i} \cup \ldots \cup U_{k_i i}$ блоков из последовательных единиц, разделенных нулями.

Алгоритм General. Алгоритм состоит из трех этапов. На первом этапе решается линейная релаксация (1)–(3). Пусть x — оптимальное решение задачи (1)–(3).

На втором этапе по решению x строится такая булева матрица A', что $d(A') = 1$. Осуществляется это следующим образом. Ограничение

(2) может быть записано в виде

$$\sum_{t=1}^{k_i} \sum_{j \in U_{it}} x_j \geqslant 1.$$

Из этого представления видно, что для некоторого блока $U_{i't}$

$$\sum_{j \in U_{it'}} x_j \geqslant 1/k_i \geqslant 1/d(A).$$

В новой матрице A' оставляем только этот блок и так поступаем со всеми строками. В итоге для каждого $i \in I$ имеем

$$\sum_{j \in J} a'_{ij} x_j \geqslant 1/d(A).$$

На третьем этапе решается задача о покрытии с матрицей A' и той же весовой функцией w. Результатом работы алгоритма General будет x' — оптимальное решение этой задачи.

Анализ алгоритма. Сложность алгоритма определяется трудоемкостью решения линейной релаксации (1)–(3).

Поскольку из описания алгоритма ясно, что матрица A' и найденное решение x', удовлетворяют условиям леммы 1 с $\alpha(A) = d(A)$, построенный алгоритм имеет оценку точности $d(A)$.

3. Алгоритм для частного случая

В этом разделе мы покажем, что применяя тот же метод можно получать алгоритмы с лучшими оценками точности для некоторых частных случаев задачи Set Cover.

Рассмотрим частный случай задачи Set Cover, в котором матрица ограничений A разбивается по строкам на r полностью унимодулярных матриц, т. е. $A = (B_1|B_2|\dots|B_r)$, где B_l — полностью унимодулярная матрица для каждого $l = 1, \dots, r$. Назовем матрицы B_l *составляющими матрицами* матрицы A. Покажем, что в этом случае задача Set Cover может быть решена полиномиальным приближенным алгоритмом с оценкой точности r.

Алгоритм Special. Алгоритм состоит из трех этапов.

На первом этапе находится оптимальное решение x линейной релаксации (1)–(3).

На втором этапе в соответствии с этим решением матрица A преобразуется в булеву матрицу $A' = (a'_{ij})$ с тем же числом строк и столбцов.

Через J_t обозначим множество столбцов матрицы A, являющихся столбцами матрицы B_t. Представим i-е ограничение из (2) в следующем виде

$$\sum_{t=1}^{r} \sum_{j \in J_t} a_{ij} x_j \geqslant 1.$$

Из этой записи ясно, что для каждого $i \in I$ можно найти число t_i такое, что

$$\sum_{j \in J_{t_i}} a_{ij} x_j \geqslant \frac{1}{r} \ .$$

Для каждого $i \in I$ и каждого $j \in J$ полагаем

$$a'_{ij} = \begin{cases} a_{ij}, & \text{если } j \in J_{t_i}, \\ 0 & \text{в противном случае.} \end{cases}$$

По построению матрица A' отличается от матрицы A тем , что каждая ее строка содержит целиком строку одной из составляющих подматриц B_l, причем все остальные элементы этой строки, соответствующие элементам других составляющих подматриц, равны нулю. Отсюда следует, что перестановками строк и столбцов матрица A' приводится к диагональному виду

$$\begin{pmatrix} B'_1 & & & \\ & B'_2 & & \\ & & \dots & \\ & & & B'_r \end{pmatrix},$$

где B'_l — подматрица матрицы B_l для каждого l, а все элементы вне диагональных подматриц равны нулю.

Напомним, что согласно определению матрица является полностью унимодулярной, если определитель любой из ее подматриц равен $0, \pm 1$. Таким образом, диагональные подматрицы матрицы A' полностью унимодулярны. Поэтому сама матрица A' полностью унимодулярна.

На третьем этапе находится оптимальное решение x' задачи Set Cover с матрицей ограничений A' и весами подмножеств w. Выше уже было отмечено, что эта задача сводится к решению задачи линейного программирования и поэтому может быть решена за полиномиальное время. Решение x' объявляется результатом работы алгоритма Special.

Анализ алгоритма. Трудоемкость построенного алгоритма, как и в случае алгоритма для общего случая, определяется трудоемкостью решения линейной релаксации.

Из построения и сказанного ясно, что матрица A' удовлетворяет условиям леммы 1 с $\alpha(A) = r$. Таким образом, по лемме 1 алгоритм Special имеет оценку точности r.

Построенный алгоритм также обобщает алгоритм из работы [8].

Укажем другое интересное следствие из полученного результата.

Рассмотрим задачу Set Cover, в которой матрица ограничений A имеет вид $A = (A_1|A_2|\ldots|A_s)$, где A_l $(l = 1,\ldots,s)$— булева матрица, содержащая в каждой строке не более одной единицы. Заметим, что при $s \leqslant 2$ матрица A полностью унимодулярна [12] и, следовательно, задача полиномиально разрешима (сводится к задаче о минимальном разрезе в двудольном графе [1, 2]). При любом $s \geqslant 3$ задача NP-трудна [2].

Заметим, что алгоритм [4] и алгоритм, построенный в предыдущем разделе, применительно к данной задаче находят решения с оценкой точности не лучше s. Из полученного в этом разделе результата следует, что данная задача может быть решена с точностью $r = \lceil s/2 \rceil$.

В самом деле, поскольку при любом $l = 1,\ldots,s-1$ матрицы A_l, $(A_l|A_{l+1})$ полностью унимодулярны, матрица A разбивается по строкам на $r = \lceil s/2 \rceil$ полностью унимодулярных подматриц:

$$A = (B_1|B_2|\ldots|B_r),$$

где $B_t = (A_{2t-1}|A_{2t})$ для $t = 1,\ldots,r-1$ и

$$B_r = \begin{cases} (A_{2r-1}|A_{2r}), & \text{если } s \text{ четно;} \\ A_{2r-1}, & \text{если } s \text{ нечетно.} \end{cases}$$

ЛИТЕРАТУРА

1. **Агеев А. А.** О минимизации некоторых полиномов от булевых переменных // Управляемые системы: Сб. науч. тр. Вып. 21. Новосибирск: Институт математики СО РАН СССР, 1981. С. 3–5.
2. **Агеев А. А.** О сложности задач минимизации полиномов от булевых переменных // Управляемые системы: Сб. науч. тр. Вып. 23. Новосибирск: Институт математики СО РАН СССР, 1983. С. 3–11.
3. **Береснев В. Л.** Алгоритмы минимизации полиномов от булевых переменных // Проблемы кибернетики. Вып. 36. М.: Наука, 1979. С. 225–246.
4. **Bar-Yehuda R., Even S.** A linear-time approximation algorithm for the weighted vertex cover problem // J. Algorithms. 1981. V. 2, N 2. P. 198–203.
5. **Chvatal V.** A greedy heuristic for the set covering problem // Math. Oper. Res. 1979. V. 4, N 3. P. 233–235.

6. **Feige U.** A threshold of $\ln n$ for approximating set cover // J. Assoc. Comput. Mach. 1998. V. 45, N 4. P. 634–652.
7. **Garey M. R., Johnson D. S.** Computers and intractability. A guide to the theory of NP-completeness. San Francisco: W. H. Freeman and Co., 1979. (Русскиу перевод: **Гэри М.,Джолнстон Д.** Вычислительные машины и трудноpешаемые задачи. М.: Мир, 1982.)
8. **Gaur D. R. , Ibaraki T., Krishnamurti R.** Constant ratio approximation algorithms for the rectangle stabbing problem and the rectilinear partitioning problem // J. Algorithms. 2002. V. 43, N 1. P. 138–152.
9. **Hassin R., Megiddo M.** Approximation algorithms for hitting objects with straight lines // Discrete Appl. Math. 1991. V. 30, N 1. P. 29–49.
10. **Heller I., Hoffman A. J.** On unimodular matrices // Pacific J. Math. 1962. V. 12, N 4. P. 1321–1327.
11. **Hochbaum D. S.** Approximation algorithms for the set covering and vertex cover problems // SIAM J. Comp. 1982. V. 11, N 3. P. 555–556.
12. **Hoffman A. J., Kruskal J. B.** Integral boundary points of convex polyhedra // Linear inequalities and related systems. Princeton: Princeton University Press, 1956. P. 223–246.
13. **Nemhauser G., Wolsey L.** A integer and combinatorial optimization. New York: John Wiley and Sons, Inc., 1988.
14. **Srinivasan A.** Improved approximations of packing and covering problems // Proc. of the 27th annual ACM symposium on theory of computing. New York: ACM, 1995. P. 268–276.

Адрес автора:

Институт математики
им. С. Л. Соболева СО РАН,
пр. Академика Коптюга, 4,
630090 Новосибирск,
Россия.
E-mail: ageev@math.nsc.ru

Статья поступила
5 января 2004 г.