# ОБ ОДНОЙ ЗАДАЧЕ ВЫБОРА ПЕРЕСТАНОВКИ СТОЛБЦОВ 0-1 МАТРИЦЫ[1]

П.А. Кононова, Ю.А. Кочетов

*Институт математики им. С.Л. Соболева СО РАН,*
*Новосибирский государственный университет, Новосибирск*
*e-mail: jkochet@math.nsc.ru*

**Аннотация.** Для задачи выбора оптимальной перестановки столбцов 0-1 матрицы предложены три формулировки задачи в терминах целочисленного линейного программирования. Используя эти формулировки, получены алгоритмы вычисления нижних и верхних оценок оптимума. Обсуждаются результаты численных экспериментов.

**Ключевые слова:** связная матрица, жадные алгоритмы, сжатие информации.

## 1. Введение

Рассмотрим следующую прикладную задачу. Кинокомпания планирует снять фильм. Он состоит из конечного числа эпизодов. В каждый день может сниматься только один эпизод, и для каждого эпизода достаточно одного дня. В эпизодах играют актеры. Актер приезжает в студию в свой первый съемочный день и вынужден оставаться там до своего последнего съемочного дня. Требуется найти такой порядок съемки эпизодов, чтобы суммарное рабочее время актеров было минимальным.

Пусть 0-1 матрица $(a_{ij})$ задает необходимый состав актеров для каждого эпизода, т.е $a_{ij} = 1$, если актер $i$ играет в эпизоде $j$ и $a_{ij} = 0$ в противном случае. Перестановка столбцов матрицы задает порядок съемки эпизодов и суммарное рабочее время каждого актера. Если существует перестановка столбцов, при которой в каждой строке единицы идут подряд, то говорят, что матрица обладает свойством связности [7]. Задача распознавания связности матрицы решается за полиномиальное время [1, 2, 6]. Если же матрица не обладает этим свойством, то задача нахождения максимальной связной подматрицы является NP-полной [1].

Задача о перестановке столбцов матрицы тесно связана с сжатием и хранением информации. Если матрица не обладает свойством связности, то нужно хранить нули, которые нарушают эту связность, т.е. окна в расписании актеров. Тогда задача сводится к поиску перестановки столбцов с минимальным числом окон. Эта задача также является NP-полной. В [3] предложен метод ветвей и границ для решения взвешенной версии этой задачи. В [9] для нее разработан генетический алгоритм. Эвристические алгоритмы исследовались в [11], однако многие из них приводят к решениям с большой погрешностью.

В настоящей работе предлагаются три формулировки задачи в терминах целочисленного линейного программирования. Они позволяют использовать коммерческое программное обеспечение для поиска оптимального решения. Эти формулировки приводят к большому разрыву целочисленности и, как следствие, позволяют решать задачи только малой размерности. Примеры с 15 эпизодами и 15 актерами даже за 24 часа машинного времени не поддаются точному решению пакетом CPLEX 11.0 на PC с процессором 2.8 ГГц. В связи с этим в работе предлагаются алгоритмы вычисления верхних и нижних оценок, использующих CPLEX для вспомогательных подзадач. Приводятся результаты численных экспериментов.

[1] Работа выполнена при финансовой поддержке РФФИ (проект 06-01-00075)

## 2. Математические формулировки

Введем следующие обозначения:

$I = \{1, \ldots, \mathbf{I}\}$ — множество актеров,

$J = \{1, \ldots, \mathbf{J}\}$ — множество эпизодов,

$T = \{1, \ldots, \mathbf{T}\}$ — множество дней для съемки фильма, $\mathbf{T} = \mathbf{J}$.

Переменные задачи:

$$x_{jt} = \begin{cases} 1, & \text{если эпизод } j \text{ снимается в день } t, \\ 0 & \text{в противном случае,} \end{cases}$$

$s_i \geq 0,$ целые — первый рабочий день актера $i$,

$f_i \geq 0,$ целые — последний рабочий день актера $i$.

С использованием введенных обозначений получаем следующую формулировку задачи в терминах целочисленного линейного программирования:

$$\min \sum_{i \in I} (f_i - s_i + 1)$$

при ограничениях:

$$\sum_{j \in J} x_{jt} = 1, \ t \in T,$$

$$\sum_{t \in T} x_{jt} = 1, \ j \in J,$$

$$f_i \geq \sum_{t \in T} t a_{ij} x_{jt}, \quad i \in I, \ j \in J,$$

$$s_i \leq \sum_{t \in T} t a_{ij} x_{jt}, \quad i \in I, \ j \in J,$$

$$f_i - s_i \geq \sum_{j \in J} a_{ij} - 1, \ \ i \in I,$$

$$x_{jt} \in \{0, 1\}, \ f_i \geq 1, \ s_i \geq 1, \ \text{целые}, \quad j \in J, t \in T, i \in I.$$

Целевая функция задачи определяет суммарное рабочее время актеров. Первое ограничение требует назначить на каждый день ровно один эпизод. Второе ограничение требует для каждого эпизода один рабочий день. Третье и четвертое ограничения определяют первый и последний рабочий день каждого актера. Пятое ограничение задает минимальные затраты на каждого из актеров. Это ограничение является вспомогательным и улучшает нижнюю оценку при релаксации целочисленных ограничений.

Полученная задача имеет $2\mathbf{I}+\mathbf{JT}$ переменных и $\mathbf{J}+\mathbf{I}+\mathbf{T}+2\mathbf{IJ}$ ограничений. Наличие больших коэффициентов в третьем и четвертом ограничениях вызывает желание ввести дополнительные переменные и представить задачу с коэффициентами 0 или 1.

Введем дополнительные переменные:

$$y_{it} = \begin{cases} 1, & \text{если актер } i \text{ находится в студии в день } t, \\ 0 & \text{в противном случае,} \end{cases}$$

$$z_{it} = \begin{cases} 1, & \text{если актер } i \text{ играет хотя бы в одном эпизоде после дня } t, \\ 0 & \text{в противном случае.} \end{cases}$$

Тогда исходная задача может быть представлена следующим образом:

$$\min \sum_{i \in I} \sum_{t \in T} y_{it}$$

при ограничениях:

$$y_{it} \geq \sum_{j \in J} a_{ij} x_{jt},\ i \in I,\ t \in T,$$

$$z_{it} \geq \sum_{j \in J} a_{ij} x_{jt'}, \quad i \in I,\ t', t \in T,\ t' > t,$$

$$y_{it} \geq y_{it'} + z_{it} - 1, \quad i \in I,\ t', t \in T,\ t' < t,$$

$$\sum_{j \in J} x_{jt} = 1,\ t \in T,$$

$$\sum_{t \in T} x_{jt} = 1,\ j \in J,$$

$$x_{jt} \in \{0, 1\},\ y_{it} \in \{0, 1\}\ z_{it} \in \{0, 1\}, \quad j \in J,\ t \in T,\ i \in I.$$

Целевая функция задачи имеет тот же смысл, что и раньше. Первое ограничение требует наличия актера в студии, если в этот день снимается эпизод с его участием. Второе ограничение показывает потребность в актере для оставшихся эпизодов. Вместе с третьим ограничением оно позволяет определить, нужно ли актеру оставаться в студии для съемок в последующие дни или эпизоды с его участием уже закончились. Остальные ограничения взяты из предыдущей модели и сохраняют свой прежний смысл. Новая формулировка имеет уже $2\mathbf{IT}+\mathbf{JT}$ переменных и $\mathbf{J}+\mathbf{T}+\mathbf{IT}+\mathbf{IT}^2$ ограничений. Ее коэффициенты равны 0 или 1. Значит для ее решения можно использовать известные отсечения задачи упаковки множества [10, 4, 5]. Следующая формулировка также обладает указанным свойством и использует переменные $y_{it}, z_{it}$, но теперь они имеют другой смысл:

$$y_{it} = \begin{cases} 1, & \text{если актер } i \text{ должен сниматься в день } t \text{ или до него,} \\ 0 & \text{в противном случае,} \end{cases}$$

$$z_{it} = \begin{cases} 1, & \text{если актер } i \text{ должен сниматься в день } t \text{ или после него,} \\ 0 & \text{в противном случае.} \end{cases}$$

Актер $i$ должен находиться в студии в день $t$, если и только если $y_{it} = z_{it} = 1$. Получаем следующую формулировку задачи:

$$\min \sum_{i \in I} \sum_{t \in T} (y_{it} + z_{it} - 1)$$

при ограничениях:

$$y_{it} \geq \sum_{j \in J} x_{jt} a_{ij}, \quad i \in I,\ t \in T,$$

$$z_{it} \geq \sum_{j \in J} x_{jt} a_{ij}, \quad i \in I,\ t \in T,$$

$$y_{it} \geq y_{it'},\ \ t', t \in T,\ t' < t,\ i \in I,$$

$$z_{it} \leq z_{it'},\ \ t', t \in T,\ t' < t,\ i \in I,$$

$$z_{it} + y_{it} \geq 1 + \sum_{j \in J} x_{jt} a_{ij}, \quad i \in I, \quad t \in T,$$

$$\sum_{j \in J} x_{jt} = 1,\ t \in T,$$

$$\sum_{t \in T} x_{jt} = 1,\ j \in J,$$

$$x_{jt} \in \{0, 1\},\ y_{it} \in \{0, 1\}\ z_{it} \in \{0, 1\}, \quad j \in J,\ t \in T,\ i \in I.$$

Целевая функция не изменила своего смысла и по–прежнему задает суммарное рабочее время актеров. Первое и второе ограничения задачи требуют присутствия актеров в день съемок эпизодов с их участием. Третье (четвертое) ограничение гарантирует неубывание (невозрастание) переменных $y_{it}$ ($z_{it}$) по времени. Пятое ограничение гарантирует, что актер получит зарплату в тот день, когда он снимается. Это ограничение является вспомогательным. Последние два ограничения перешли из предшествующих постановок. Число переменных задачи не изменилось, их по–прежнему $2\mathbf{IT}+\mathbf{JT}$. Число ограничений немного возросло, их стало $\mathbf{J}+\mathbf{T}+3\mathbf{IT}+\mathbf{IT}^2$.

### 3. Нижние оценки

Как уже упоминалось выше, использование коммерческого программного обеспечения для решения сформулированной задачи наталкивается на проблему большого разрыва целочисленности. В связи с этим возникает потребность в разработке приближенных алгоритмов и оценке качества их работы. Для получения нижних оценок оптимума можно использовать, например, оценки линейных релаксаций приведенных выше формулировок задачи. Однако они оказываются не лучше тривиальной нижней оценки $\sum_{i\in I}\sum_{j\in J} a_{ij}$. Для повышения нижней оценки можно потребовать целочисленности переменных $x_{jt}$, но не для всех $t$, что соответствовало бы оптимальному решению исходной задачи, а только для малой их части, выбранной специальным образом. В ходе численных экспериментов исследовались две стратегии:

**Стратегия 1.** Выбрать $t$, для которого в оптимальном решении $x^*_{jt}$ линейной релаксации величина $\max_{j\in J} x^*_{jt}$ была минимальной.

**Стратегия 2.** Выбрать $t$, для которого в оптимальном решении $x^*_{jt}$ линейной релаксации мощность множества $\{j \in J \mid x^*_{jt} > 0\}$ была максимальной.

Обе стратегии стараются найти столбец, где условие целочисленности нарушается в наибольшей степени. Результаты экспериментальных исследований приведены в Таблице 1. Каждая строка таблицы соответствует одному примеру размерности $\mathbf{I}=\mathbf{J}=15$. Каждый элемент матрицы $(a_{ij})$ с вероятностью 0,3 независимо от других элементов полагался равным 1. Первая часть таблицы (1 столбец) представляет результаты расчетов при условии целочисленности одного столбца по 1-й (MinMax) и 2-й (Max$\neq 0$) стратегиях, а также наилучшую нижнюю оценку (Max), получаемую последовательным перебором всех столбцов. Наряду с нижней оценкой указывается время счета для ее получения. Столбец *Lp* показывает тривиальную нижнюю оценку. Представленные результаты показывают, что выбор одного столбца по любому правилу мало влияет на величину нижней оценки. Оставшаяся часть таблицы показывает результаты при фиксации двух столбцов. Нижняя оценка возрастает. Первая стратегия показывает наихудшие результаты. Вторая стратегия приводит к результатам, мало отличающимся от стратегий брать первые два столбца (1 и 2) или первый и последний (1 и 15). По-видимому, на данном классе следует брать несколько разных пар и выбирать лучшую из найденных оценок, либо фиксировать три столбца или больше.

Таблица 1: Нижние оценки оптимума

| № | Lp | 1 столбец | | | | | 2 столбца | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MinMax | | Max$\neq 0$ | | Max | MinMax | | Max$\neq 0$ | | 1 и 15 | | 1 и 2 | |
| 1 | 65 | 66 | 0:0 | 66 | 0:0 | 66 | 67 | 0:10 | 70 | 0:10 | 70 | 0:22 | 71 | 0:54 |
| 2 | 64 | 64 | 0:0 | 67 | 0:0 | 67 | 69 | 0:18 | 72 | 0:23 | 75 | 0:40 | 72 | 1:04 |
| 3 | 68 | 68 | 0:05 | 68 | 0:02 | 68 | 70 | 1:06 | 70 | 0:53 | 70 | 1:34 | 70 | 1:05 |

## 4. Верхние оценки

Так как задача с требованием целочисленности одного столбца матрицы $(x_{jt})$ решается легко, то можно предложить следующий простой алгоритм получения допустимого решения задачи.

**Алгоритм**

1. Найти оптимальное решение $(x^*_{jt})$ линейной релаксации.
2. Выбрать столбец в матрице $(x^*_{jt})$ с дробными значениями.
3. Решить задачу с условием целочисленности для данного столбца.
4. Зафиксировать полученные значения для данного столбца, сократить размерность задачи и вернуться на шаг 2, если оставшееся число столбцов не меньше заданного порога $\mathbf{T}'$.
5. Решить точно задачу при $\mathbf{T} = \mathbf{T}'$.

Выбор параметра $\mathbf{T}'$ зависит от производительности компьютера и сложности примера. В Таблице 2 приведены результаты расчетов при различных $\mathbf{T}'$. В качестве стратегии выбора столбца на шаге 2 исследовались следующие два правила.

**Стратегия 1.** Выбрать $t$, для которого в оптимальном решении $x^*_{jt}$ линейной релаксации величина $\max_{j \in J} x^*_{jt}$ была максимальной.

**Стратегия 2.** Выбрать $t$, для которого в оптимальном решении $x^*_{jt}$ линейной релаксации мощность множества $\{j \in J \mid x^*_{jt} > 0\}$ была минимальной.

Согласно этим стратегиям выбираются столбцы в которых решение наиболее близко к целочисленному. Строки Таблицы 2 соответствуют тем же примерам, что и в Таблице 1. Столбцы MaxMax соответствуют первой стратегии. Столбцы Min$\neq 0$ соответствуют второй стратегии. Результаты расчетов свидетельствуют что при небольших $\mathbf{T}'$ лучше использовать 1-ю стратегию, а при больших $\mathbf{T}'$ применять 2-ю стратегию.

Таблица 2: Верхние оценки оптимума

| № | OPT | $\mathbf{T}' = 12$ | | | | $\mathbf{T}' = 10$ | | | | $\mathbf{T}' = 7$ | | | | $\mathbf{T}' = 5$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MaxMax | | Min$\neq$ 0 | | MaxMax | | Min$\neq$ 0 | | MaxMax | | Min$\neq$ 0 | | MaxMax | | Min$\neq$ 0 | |
| 1 | 112 | 117 | 3:04 | 117 | 3:04 | 130 | 0:24 | 122 | 0:10 | 134 | 0:05 | 132 | 0:04 | 136 | 0:0 | 137 | 0:0 |
| 2 | 101 | 112 | 1:14 | 109 | 1:14 | 120 | 0:30 | 128 | 0:15 | 146 | 0:04 | 153 | 0:05 | 153 | 0:0 | 158 | 0:0 |
| 3 | 104 | 124 | 4:10 | 123 | 4:10 | 132 | 0:16 | 135 | 0:14 | 139 | 0:06 | 149 | 0:04 | 144 | 0:0 | 155 | 0:0 |

Заметим, что время счета резко падает при фиксации даже небольшого числа столбцов. Выбор столбцов играет важную роль, так как влияет на размерность получаемой подзадачи. Следующая стратегия, несмотря на скромные показатели при вычислении нижних оценок, оказалась удивительно продуктивной.

**Стратегия 3.** Выбрать сначала минимальное значение $t$, а затем максимальное значение.

В Таблице 3 приведены результаты расчетов для этой стратегии. Малая погрешность ($<$5%) даже при $\mathbf{T}' = 7$ свидетельствует о продуктивности предложенного подхода.

Таблица 3: Верхние оценки оптимума

| № | OPT | $\mathbf{T}' = 13$ | | $\mathbf{T}' = 11$ | | $\mathbf{T}' = 9$ | | $\mathbf{T}' = 7$ | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 112 | 112 | 40:19 | 112 | 1:23 | 114 | 0:11 | 118 | 0:10 |
| 2 | 101 | 101 | 11:22 | 105 | 0:36 | 105 | 0:30 | 105 | 0:30 |
| 3 | 104 | 104 | 3:13 | 104 | 0:19 | 106 | 0:09 | 107 | 0:09 |

Авторы выражают искреннюю благодарность Антону Валентиновичу Еремееву за предоставленную возможность проведения вычислительных экспериментов в алгебраической моделирующей системе GAMS.

# Список литературы

[1] K.S. Booth *PQ-tree algorithms* - PH.D. thesis, University of California, Berkeley, 1975.

[2] K.S. Booth and G. S. Lueker *Testing for the consecutive ones property, inter- vals graphs, and graph planarity using PQ-tree algorithm* - J.Comput.System.Sci., 1976, v. 13, p. 335–379.

[3] T.C.E. Cheng, J.E. Diamond, B.M.T. Lin *Optimal scheduling in .lm production to minimize talent hold cost* - Journal of Optimization Theory and Applications, 1993, v. 79 N. 3, p. 479–492.

[4] E. Cheng and W.Y. Cunninghav. *Wheel inequalities for stable set polytopes* - Mathematical Programming, 1997, v.77, N 3, p. 389–421.

[5] E. Cheng and S. Vries. *Antiweb-wheel inequalities and their separation problems over the stable set polytopes* - Mathematical Programming, 2002, v.92, N 1, p. 153–175.

[6] D.R. Fulkerson and O.A. *Gross, Incidence matrices and interval graphs* - Pacif. J. Math, 1965, v. 15, p. 835–855.

[7] M.R. Garey and D.S. Johnson *Computers and Intractability: A Guide to the Theory of NP-Completeness* - Freedman, 1979, San Francisco, CA.

[8] W.L. Hsu *A simple test for the consecutive ones property* - Journal of Algorithms, 2002, v. 43, p. 1–16.

[9] A.L. NordstrAom, S. Tufekci *A genetic algorithm for the talent scheduling problem* - Computers and Operations Research, 1994, v. 21, N. 8, p. 927–940.

[10] F. Rossi and S. Smriglio. *A branch-and-cut algorithm for the maximum cardinality stable set problem* - Operations Research Letters, 2001, v. 28, p. 63–74.

[11] M. Veldhorst *Approximation of the consecutive ones matrix augmentation problem* - SIAM Journal on Computing, 1985, v. 14, p. 709–729.

# ON A PROBLEM OF PERMUTATION OF THE COLUMNS FOR 0-1 MATRIX

P. Kononova, Yu. Kochetov

*Sobolev Institute of Mathematics, Novosibirsk State University, Novosibirsk*
*e-mail: jkochet@math.nsc.ru*

**Abstract.** We consider three linear integer programming formulations for the problem of permutation of the columns for 0-1 matrix. Using these formulations we present some upper and lower bounds on optimal solutions. Computational results are discussed.

**Key words:** connected matrices, greedy algorithms, data compression.