УДК 519.812.5

О. К. ИЛЮНИН, канд. техн. наук, *Б. В. НОВИКОВ*

О НАХОЖДЕНИИ МЕДИАНЫ ДЛЯ ЭКСПЕРТНЫХ УПОРЯДОЧЕНИЙ

Пусть $M = \{1, \ldots, m\}$ — множество сравниваемых объектов, $R_1, \ldots, R_n$ — заданные на этом множестве линейные упорядочения.

Определение. Упорядоченной дихотомией на множестве M назовем такое отношение $D \subseteq M \times M$, что $D = (A \times M) \cup ((M \setminus A) \times (M \setminus A))$ для некоторого подмножества $A \subseteq M$. Такое отношение называют дихотомическим линейным квазипорядком [1].

Очевидно, что упорядоченная дихотомия является линейным квазипорядком, разбивающим множество M на две группы объектов («хорошие» и «плохие» объекты). Матрица, соответствующая упорядоченной дихотомии, является блочно-треугольной и имеет вид (при подходящем упорядочении объектов) $\left(\begin{array}{c|c} 1 & 1 \\ \hline 0 & 1 \end{array}\right)$, где разбиение на блоки соответствует разбиению множества M на подмножества A и $M \setminus A$.

Если трактовать бинарные отношения $R_1, \ldots, R_n$ как индивидуальные упорядочения экспертов, возникает задача о нахождении группового транзитивного отношения, т. е. отношения наиболее приемлемого в некотором смысле для всех экспертов., Один из способов ее решения состоит в следующем [1]. В классе всех транзитивных отношений T находят минимум суммы

$$\sum_{i=1}^{n} d(R_i, R), \tag{1}$$

где R пробегает множество T; $d(R_i, R) = |(R_i \cup R) \setminus (R_i \cap R)|$ — расстояние Хемминга. Отношение $R \in T$, минимизирующее выражение (1), будем в дальнейшем называть медианой отношений $R_1, \ldots, R_n$ в классе T [1]. Как известно, отыскание минимума (1) сводится к задаче целочисленного программирования и не имеет в настоящее время достаточно эффективного решения в классе всех транзитивных отношений. В связи с этим возникает вопрос: для каких подмножеств $T' \subseteq T$ можно указать эффективный алгоритм нахождения минимума (1) в множестве T'? Как показывают выкладки в [1], такой алгоритм вряд ли возможно найти в случаях, когда T' — множество всех эквивалентностей или квазипорядков.

В данном сообщении предлагается способ нахождения медианы линейных порядков в множестве T_{0d} упорядоченных дихотомий.

Предложенный алгоритм является достаточно простым и не сводится к задаче целочисленного программирования.

Пусть R — линейный порядок, D — упорядоченная дихотомия. Удобно представлять отношения в виде матриц размера $m \times m$ из нулей и единиц. Например, $R = (r_{ij})$, где

$$r_{ij} = \begin{cases} 1, & \text{если } (i, j) \in R; \\ 0, & \text{если } (i, j) \notin R. \end{cases}$$

Аналогично будем писать $D = (d_{ij})$. Введем обозначения $\bar{A} = M \setminus A$ и $a = |A|$ (A — подмножество, соответствующее упорядоченной дихотомии D, см. определение).

Вычислим расстояние между R и D. Заметим предварительно, что $|D| = |A||M| + (|M| - |A|)^2 = m^2 - ma + a^2$ и $|R| = \frac{m(m+1)}{2}$. Кроме того, поскольку $R \cap (\bar{A} \times \bar{A})$ — линейный порядок на множестве $\bar{A}$, то $|R \cap (\bar{A} \times \bar{A})| = \frac{(m-a)(m-a+1)}{2}$. Поэтому

$$d(R, D) = |R| + |D| - 2|R \cap D| = |R| + |D| - 2|R \cap (\bar{A} \times \bar{A})| - 2|R \cap (A \times M)| = \frac{m(m+1)}{2} + m^2 - ma + a^2 - (m - a)(m + 1 - a) - 2 \sum_{i \in A} \sum_{j=1}^{m} r_{ij} = \frac{m(m-1)}{2} + (m+1)a - 2 \sum_{i \in A} r_i,$$

где через r_{ij} обозначена сумма Σr_{ij}. Таким образом, r_i — ранг объекта i в упорядочении R. Минимизацию суммарного расстояния можно разбить на два этапа: зафиксировав число $a = |A|$, найдем

$$\Delta_a = \min_{|A|=a} \sum_{i=1}^{n} d(R_i, D),$$

а затем отыщем $\min\limits_{0 \leqslant a \leqslant m} \Delta_a$.

Согласно предыдущим выкладкам Δ_a имеет вид

$$\Delta_a = \frac{mn(m-1)}{2} + (m+1)an - 2 \max_{A:|A|=a} \sum_{i \in A} S_i,$$

где $S_i = \sum\limits_{j=1}^{n} r_i^j$. Перенумеруем теперь элементы множества M так, чтобы упорядочить суммы S_i: $S_1 \geqslant S_2 \geqslant \ldots \geqslant S_m$. Тогда нетрудно увидеть, что

$$\Delta_a = \frac{mn(m-1)}{2} + (m+1)an - 2 \sum_{i=1}^{a} S_i. \tag{2}$$

Тем самым первый этап минимизации завершен. Попутно решена следующая задача: найти медиану линейных порядков $R_1, \ldots, R_n$ в классе T^a_{od} упорядоченных дихотомий с фиксированной мощностью множества A, $a = |A|$. Для этого нужно найти все суммы S_i и, упорядочив соответственно элементы множества M, взять в качестве множества A первые a элементов. В частности, при $a = 1$ из (2) получаем

$$\Delta_1 = \frac{mn(m-1)}{2} + (m+1)n - 2S_1 = \frac{n(m^2+m+2)}{2} - 2S_1.$$

Таким образом, «наилучшим» из элементов множества M будет тот элемент x, для которого максимальна сумма

$$S_x = \sum_{i=1}^{m} \sum_{j=1}^{n} r^j_{xi}.$$

Находя последовательно «наилучший» элемент y в множестве $M \setminus \{x\}$, затем «наилучший» элемент z в множестве $M \setminus \{x, y\}$ и т. д., мы получаем упорядочение элементов $x \geqslant y \geqslant z \geqslant \ldots$, которое соответствует упорядочению, полученному суммированием рангов, приписанных каждому объекту экспертами [2].

Для решения второго этапа задачи рассмотрим разность

$$\Delta_a - \Delta_{a-1} = (m+1)n - 2S_a. \tag{3}$$

Из (3) следует, что $\Delta_a \leqslant \Delta_{a-1}$, и только тогда, когда $S_a \geqslant \frac{(m+1)n}{2}$. Поскольку величины S_α $(\alpha \in M)$ упорядочены, то функция $f(a) = \Delta_a$ имеет, следовательно, минимум на отрезке $[0, m]$, который достигается в точке a, такой, что

$$S_a \geqslant \frac{(m+1)n}{2} \geqslant S_{a+1}.$$

Точка минимума может быть не единственной, так как возможно существование отрезка, содержащего a, во всех точках которого достигается минимум.

Таким образом, для нахождения искомой упорядоченной дихотомии D достаточно после нахождения сумм S_α взять в качестве A подмножество всех элементов $\alpha \in M$, таких, что $S_\alpha \geqslant \frac{(m+1)n}{2}$.

Пример. Пусть 4 эксперта определяют следующие упорядочения множества объектов $M = \{1, 2, 3, 4, 5\}$: R_1: $2 > 3 > 1 > 4 > 5$; R_2: $4 > 5 > 1 > 3 > 2$; R_3: $3 > 2 > 5 > 4 > 1$; R_4: $5 > 1 > 3 > 4 > 2$. Тогда $S_1 = 11$; $S_2 = 11$; $S_3 = 14$; $S_4 = 11$; $S_5 = 13$. Поскольку $(m+1)n/2 = 12$, то $A = \{3, 5\}$ — подмножество «лучших» объектов, а $\bar{A} = \{1, 2, 4\}$ — подмножество «худших» объектов.

В заключение заметим, что предложенный алгоритм может быть использован для выделения группы значимых признаков из множества признаков, характеризующих некоторый объект, если эти признаки предварительно упорядочены экспертами по значимости. Задачи такого типа возникают при использовании моделей анализа неколичественной информации, шкалировании, классификации и т. п.

Список литературы: **1.** *Миркин Б. Г.* Проблема группового выбора. М., Наука, 1974. 400 с. **2.** *Кендэл М.* Ранговые корреляции. М., Статистика, 1975. 200 с.

Поступила 13 октября 1978 г.