ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь—февраль 2011. Том 18, № 1. С. 61–69

УДК 519.2+621.391

ПРИБЛИЖЁННЫЙ АЛГОРИТМ РЕШЕНИЯ ОДНОЙ ЗАДАЧИ ПОИСКА ПОДМНОЖЕСТВА ВЕКТОРОВ *)

А. В. Кельманов, С. М. Романченко

Аннотация. Одна из проблем анализа данных сводится к решению NP-трудной экстремальной задачи поиска в множестве векторов евклидова пространства подмножества векторов, имеющего заданную мощность и включающего векторы, «близкие» между собой по критерию минимума суммы квадратов расстояний. В работе обоснован полиномиальный 2-приближённый алгоритм решения этой задачи.

Ключевые слова: поиск подмножества векторов, NP-трудность, эффективный приближённый алгоритм.

Введение

Объектом исследования настоящей работы являются проблемы оптимизации в задачах анализа данных, предметом исследования — труднорешаемая экстремальная задача, к которой сводится одна из проблем поиска подмножества векторов евклидова пространства. Цель работы — обоснование приближённого алгоритма решения этой задачи.

Содержательная проблема анализа данных состоит в следующем. Имеется таблица, содержащая результаты измерения набора числовых информационно значимых характеристик для совокупности некоторых материальных объектов. Часть объектов из этой совокупности идентичны и имеют одинаковые характеристики. Число идентичных объектов известно. Оставшиеся объекты различны и имеют отличающиеся характеристики. В каждом результате измерения, представленном в таблице, имеется ошибка, причём соответствие между объектом и набором неизвестно. Требуется, используя адекватный измеряемым характеристикам

*) Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проекты 09–01–00032, 10–07–00195), целевой программы АВЦП Рособразования (проект № 2.1.1/3235), федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» (гос. контракт № 14.740.11.0362), целевой программы № 2 Президиума РАН (проект № 227), а также целевой программы СО РАН (интеграционный проект № 44).

критерий, найти подмножество наборов, соответствующих идентичным объектам, и оценить по результатам измерения набор характеристик этих объектов (учитывая, что данные содержат ошибку измерения).

В [1] дана формулировка этой содержательной проблемы как задачи поиска в множестве векторов евклидова пространства такого подмножества векторов фиксированной мощности, что его элементы «близки» по критерию минимума суммы квадратов расстояний. Там же установлено, что её решение сводится к решению следующей NP-трудной задачи.

Задача VS–2 (Vector Subset 2). *Дано:* множество $\mathcal{Y} = \{y_1, \ldots, y_N\}$ векторов из $\mathbb{R}^q$ и натуральное число $M > 1$. *Найти:* подмножество $\mathcal{C} \subseteq \mathcal{Y}$ векторов такое, что целевая функция

$$F(\mathcal{C}) = \sum_{y \in \mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2, \qquad (1)$$

где $\overline{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum\limits_{y \in \mathcal{C}} y$, минимальна при ограничении $|\mathcal{C}| = M$ на мощность искомого подмножества.

Мотивацией исследований послужил тот факт, что какие-либо эффективные алгоритмы с гарантированными оценками точности решения этой задачи были неизвестны. Ниже обоснован эффективный 2-приближённый алгоритм её решения.

1. Приближённый алгоритм решения задачи

Обозначим через $\mathcal{C}^*$ оптимальное решение задачи VS–2. Положим $c^* = \overline{y}(\mathcal{C}^*)$, где $\overline{y}(\cdot)$ — функция, определённая в формулировке этой задачи. Свойство оптимального решения устанавливает простая

Лемма 1. *Для любых векторов $y \in \mathcal{C}^*$ и $z \in \mathcal{Y} \setminus \mathcal{C}^*$ справедливо неравенство*

$$\|y - c^*\| \leqslant \|z - c^*\|.$$

Доказательство. Допустим противное, т. е. что существуют векторы $u \in \mathcal{C}^*$ и $v \in \mathcal{Y} \setminus \mathcal{C}^*$ такие, что $\|u - c^*\| > \|v - c^*\|$. Тогда

$$\begin{aligned} F(\mathcal{C}^*) = \sum_{y \in \mathcal{C}^*} \|y - c^*\|^2 &= \sum_{y \in \mathcal{C}^* \setminus \{u\}} \|y - c^*\|^2 + \|u - c^*\|^2 \\ &> \sum_{y \in \mathcal{C}^* \setminus \{u\}} \|y - c^*\|^2 + \|v - c^*\|^2 \\ = \sum_{y \in \{\mathcal{C}^* \setminus \{u\}\} \cup \{v\}} \|y - c^*\|^2 &\geqslant \sum_{y \in \{\mathcal{C}^* \setminus \{u\}\} \cup \{v\}} \|y - \overline{y}(\{\mathcal{C}^* \setminus \{u\}\} \cup \{v\})\|^2 \end{aligned}$$

$$= F(\{\mathcal{C}^* \setminus \{u\}\} \cup \{v\}),$$

что противоречит оптимальности множества $\mathcal{C}^*$. Лемма 1 доказана.

Лемма 1 показывает, что оптимальное решение задачи VS–2 — подмножество $\mathcal{C}^* \subseteq \mathcal{Y}$ — состоит из векторов, ближайших к вектору c^* по расстоянию. Она устанавливает необходимое условие минимума и указывает на изложенный ниже возможный подход к решению задачи.

Рассмотрим следующую вспомогательную задачу.

Задача VSVN (Vector and Subset of Vectors which are Nearest to this vector). *Дано:* множество $\mathcal{Y} = \{y_1, \ldots, y_N\}$ векторов из $\mathbb{R}^q$ и натуральное число $M > 1$. *Найти:* подмножество $\mathcal{B} \subseteq \mathcal{Y}$ векторов мощности M и вектор $b \in \mathcal{Y}$ такие, что целевая функция

$$G(\mathcal{B}, b) = \sum_{y \in \mathcal{B}} \|y - b\|^2 \tag{2}$$

минимальна.

Построим алгоритм решения этой задачи. Обозначим через $\mathcal{B}^*$ и b^* подмножество и вектор, доставляющие минимум G^* целевой функции G.

Алгоритм $\mathcal{A}_1$.

Шаг 1. Для каждого вектора $y \in \mathcal{Y}$ найдём множество $\mathcal{B}(y)$, состоящее из вектора y и $M-1$ векторов множества $\mathcal{Y}$, ближайших (по расстоянию) к вектору y. Вычислим значение целевой функции $G(\mathcal{B}(y), y)$.

Шаг 2. Среди найденных на шаге 1 множеств выберем в качестве решения множество $\mathcal{B}^*$ и вектор b^*, для которых значение целевой функции G минимально. Если минимальному значению целевой функции соответствует несколько решений, то в качестве окончательного решения выберем любое из них.

Оценку сложности и точности алгоритма устанавливает

Лемма 2. *Алгоритм $\mathcal{A}_1$ находит оптимальное решение задачи* VSVN *за время $\mathcal{O}(qN^2)$.*

Доказательство. Оптимальность решения следует из определения подмножества $\mathcal{B}(y) \subseteq \mathcal{Y}$, $y \in \mathcal{Y}$, (см. шаг 1 алгоритма $\mathcal{A}_1$) и очевидной цепочки равенств

$$G^* = G(\mathcal{B}^*, b^*) = \min_{y \in \mathcal{Y},\, \mathcal{B} \subseteq \mathcal{Y}} G(\mathcal{B}, y) = \min_{y \in \mathcal{Y}} \min_{\mathcal{B} \subseteq \mathcal{Y}} G(\mathcal{B}, y) = \min_{y \in \mathcal{Y}} G(\mathcal{B}(y), y).$$

Оценим временну́ю сложность алгоритма. На шаге 1 для отыскания множества $\mathcal{B}(y)$ необходимо вычислить элементы массива расстояний от

вектора y до каждого вектора из $\mathcal{Y}$, что потребует $\mathcal{O}(qN)$ операций, так как размерность пространства равна q, а $|\mathcal{Y}| = N$. Затраты на формирование множества $\mathcal{B}(y)$ составят $\mathcal{O}(N)$ операций (например, с помощью алгоритма [3] поиска M-го наименьшего числа в массиве из N чисел). Для вычисления значения $G(\mathcal{B}(y), y)$ по сформированному множеству $\mathcal{B}(y)$ и вектору y необходимо $\mathcal{O}(M)$ операций, что не превышает $\mathcal{O}(N)$. Суммируя перечисленные затраты, устанавливаем, что потребуется $\mathcal{O}(qN)$ операций для вычисления значения целевой функции при каждом $y \in \mathcal{Y}$.

Поскольку вычисление значения целевой функции выполняется для каждого $y \in \mathcal{Y}$, а выбор наименьшего значения целевой функции в множестве вычисленных значений на шаге 2 потребует не более $\mathcal{O}(N)$ операций сравнения, итоговая временна́я сложность равна $\mathcal{O}(qN^2)$. Лемма 2 доказана.

Докажем вспомогательное утверждение.

Лемма 3. *Пусть $\mathcal{Z}$ — непустое конечное множество векторов из $\mathbb{R}^q$ и $\overline{z}(\mathcal{Z}) = \frac{1}{|\mathcal{Z}|} \sum\limits_{z \in \mathcal{Z}} z$. Тогда если вектор $x \in \mathbb{R}^q$ удовлетворяет условиям*

$$\|x - \overline{z}\| \leqslant \|z - \overline{z}\| \qquad \forall z \in \mathcal{Z}, \tag{3}$$

то имеет место неравенство

$$\sum_{z \in \mathcal{Z}} \|z - x\|^2 \leqslant 2 \sum_{z \in \mathcal{Z}} \|z - \overline{z}\|^2. \tag{4}$$

Доказательство. Раскрывая сумму квадратов в левой части неравенства (4) с учётом (3), получим

$$\begin{aligned}
\sum_{z \in \mathcal{Z}} \|z - x\|^2 &= \sum_{z \in \mathcal{Z}} \|z - \overline{z} + \overline{z} - x\|^2 \\
&= \sum_{z \in \mathcal{Z}} \|z - \overline{z}\|^2 + |\mathcal{Z}| \cdot \|\overline{z} - x\|^2 + 2 \sum_{z \in \mathcal{Z}} \langle z - \overline{z}, \overline{z} - x \rangle \\
&= \sum_{z \in \mathcal{Z}} \|z - \overline{z}\|^2 + |\mathcal{Z}| \cdot \|\overline{z} - x\|^2 + 2 \Big\langle \sum_{z \in \mathcal{Z}} z - |\mathcal{Z}| \overline{z}, \overline{z} - x \Big\rangle \\
&= \sum_{z \in \mathcal{Z}} \|z - \overline{z}\|^2 + |\mathcal{Z}| \cdot \|\overline{z} - x\|^2 \leqslant 2 \sum_{z \in \mathcal{Z}} \|z - \overline{z}\|^2,
\end{aligned}$$

где $\langle \cdot, \cdot \rangle$ — скалярное произведение векторов. Лемма 3 доказана.

Лемма 4. *Пусть $\mathcal{B}^*$, b^* — оптимальное решение задачи* VSVN, *$\mathcal{C}^*$ — оптимальное решение задачи* VS–2. *Тогда $F(\mathcal{B}^*) \leqslant 2F(\mathcal{C}^*)$.*

Доказательство. Для оптимального решения $\mathcal{C}^*$ задачи VS–2 имеем вектор $y^* = \overline{y}(\mathcal{C}^*) = \frac{1}{|\mathcal{C}^*|}\sum\limits_{y\in\mathcal{C}^*} y$. Рассмотрим вектор $t = \arg\min\limits_{y\in\mathcal{C}^*} \|y - y^*\|$. Этот ближайший к y^* вектор в множестве $\mathcal{C}^*$ и само множество $\mathcal{C}^*$ удовлетворяют условиям леммы 3. Поэтому

$$\sum_{y\in\mathcal{C}^*} \|y - t\|^2 \leqslant 2\sum_{y\in\mathcal{C}^*} \|y - y^*\|^2 = 2F(\mathcal{C}^*). \tag{5}$$

С другой стороны, $|\mathcal{C}^*| = |\mathcal{B}^*| = M$, $t \in \mathcal{Y}$ и $\mathcal{C}^* \subseteq \mathcal{Y}$. Поэтому $\mathcal{C}^*$, t — допустимое решение вспомогательной задачи VSVN. Следовательно, для левой части (5) справедлива оценка

$$\sum_{y\in\mathcal{C}^*} \|y - t\|^2 = G(\mathcal{C}^*, t) \geqslant G(\mathcal{B}^*, b^*). \tag{6}$$

Кроме того, легко проверить, что для любого непустого конечного множества $\mathcal{Z}$ векторов из $\mathbb{R}^q$ минимум суммы квадратов $\sum\limits_{z\in\mathcal{Z}} \|z - x\|^2$ по x доставляется вектором $x^* = \frac{1}{|\mathcal{Z}|}\sum\limits_{z\in\mathcal{Z}} z = \overline{z}(\mathcal{Z})$. Поэтому для множества $\mathcal{B}^*$, векторов b^* и $\overline{y}(\mathcal{B}^*) = \frac{1}{|\mathcal{B}^*|}\sum\limits_{y\in\mathcal{B}^*} y$ имеем неравенство

$$\sum_{y\in\mathcal{B}^*} \|y - b^*\|^2 \geqslant \sum_{y\in\mathcal{B}^*} \|y - \overline{y}(\mathcal{B}^*)\|^2.$$

Используя это неравенство и определение (2), для правой части (6) получим оценку

$$G(\mathcal{B}^*, b^*) = \sum_{y\in\mathcal{B}^*} \|y - b^*\|^2 \geqslant \sum_{y\in\mathcal{B}^*} \|y - \overline{y}(\mathcal{B}^*)\|^2 = F(\mathcal{B}^*). \tag{7}$$

Объединяя (5)–(7), получим цепочку оценочных неравенств

$$F(\mathcal{B}^*) \leqslant G(\mathcal{B}^*, b^*) \leqslant G(\mathcal{C}^*, t) \leqslant 2F(\mathcal{C}^*).$$

Лемма 4 доказана.

Опираясь на лемму 4, представим алгоритм решения задачи VS–2.

Алгоритм $\mathcal{A}$.

Шаг 1. По заданному множеству $\mathcal{Y}$ и числу M находим оптимальное решение $\mathcal{B}^*, b^*$ вспомогательной задачи VSVN с помощью алгоритма $\mathcal{A}_1$.

Шаг 2. Подмножество $\mathcal{B}^*$ объявляем решением задачи VS–2.

Теорема 1. *Алгоритм $\mathcal{A}$ находит приближённое решение задачи* VS–2 *с гарантированной оценкой точности 2 за время $\mathcal{O}(qN^2)$. Оценка 2 точности алгоритма достижима и неулучшаема.*

Доказательство. Справедливость утверждения теоремы следует из лемм 1–4 и приведённого ниже примера, показывающего существование таких входных данных задачи, для которых отношение $F(\mathcal{B}^*)/F(\mathcal{C}^*)$ может быть сколь угодно близко к 2 и равно 2.

Пример. Пусть $q = 2$, $N = 4$, $M = 3$, $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$, где $y_1 = (0,0)$, $y_2 = (1,0)$, $y_3 = (-1,0)$, $y_4 = (1/2, \alpha/2)$ и $\sqrt{3} < \alpha < 3$.

Легко видеть, что оптимальным решением задачи VS–2 в этом примере является подмножество $\mathcal{C}^* = \{y_1, y_2, y_4\}$, причём $F(\mathcal{C}^*) = 1 + \frac{\alpha^2-3}{6}$. Для алгоритмического решения имеем $\mathcal{B}^* = \{y_1, y_2, y_3\}$, $F(\mathcal{B}^*) = 2$. Отношение $F(\mathcal{B}^*)/F(\mathcal{C}^*) = 2/\big(1 + \frac{\alpha^2-3}{6}\big)$ может быть сколь угодно близко к 2 при $\alpha \to \sqrt{3}$, что означает неулучшаемость полученной оценки для построенного алгоритма.

Как известно, задание иррациональных чисел на входе задачи осуществимо лишь с погрешностью. Поэтому в данном примере равенство $\alpha = \sqrt{3}$ при задании второй координаты вектора y_4 осуществимо лишь теоретически. Тем не менее, очевидно, что из-за погрешности в представлении числовых данных в компьютере допустимым является случай, когда в процессе вычислений будет иметь место равенство $\alpha^2 = 3$. В этом случае оптимальным решением задачи является подмножество $\mathcal{C}^* = \{y_1, y_2, y_4\}$, причём $F(\mathcal{C}^*) = 1$. Возможны два равноправных алгоритмических решения $\mathcal{B}_1^* = \{y_1, y_2, y_3\}$ и $\mathcal{B}_2^* = \{y_1, y_2, y_4\}$, так как $G(\mathcal{B}_1^*, y_1) = G(\mathcal{B}_2^*, y_1) = 2$, причём $F(\mathcal{B}_1^*) = 2$, а $F(\mathcal{B}_2^*) = 1$. Для первого решения имеет место равенство $F(\mathcal{B}_1^*)/F(\mathcal{C}^*) = 2$, т. е. оценка 2 точности алгоритма достижима. Теорема 1 доказана.

2. Применение алгоритма к решению других задач

В [1] показано, что задача VS–2 эквивалентна следующей NP-трудной задаче.

Задача MSSC–Case. *Дано*: множество $\mathcal{Y} = \{y_1, \ldots, y_N\}$ векторов из $\mathbb{R}^q$ и натуральное число $M > 1$. *Найти*: разбиение множества $\mathcal{Y}$ на $N - M + 1$ непустых кластеров $\mathcal{C}_1, \ldots, \mathcal{C}_{N-M+1}$ такое, что мощность одного из этих кластеров равна M и

$$R(\mathcal{C}_1, \ldots, \mathcal{C}_{N-M+1}) = \sum_{j=1}^{N-M+1} \sum_{y \in \mathcal{C}_j} \|y - \overline{y}_j(\mathcal{C}_j)\|^2 \to \min,$$

где $\overline{y}_j(\mathcal{C}_j) = \frac{1}{|\mathcal{C}_j|} \sum\limits_{y \in \mathcal{C}_j} y$, $j = 1, \ldots, N - M + 1$, — центр j-го кластера.

Эту задачу можно интерпретировать как разновидность задачи кластерного анализа данных. Для отыскания её приближённого решения можно применить алгоритм $\mathcal{A}$.

Действительно, пусть, например, мощность кластера $\mathcal{C}_1$ равна M. Тогда из условий задачи следует, что мощность оставшихся $N - M$ кластеров равна 1. Поскольку центр кластера, имеющего мощность, равную 1, совпадает с единственным вектором этого кластера, для целевой функции задачи MSSC–Case имеем

$$R(\mathcal{C}_1, \ldots, \mathcal{C}_{N-M+1}) = \sum_{j=1}^{N-M+1} \sum_{y \in \mathcal{C}_j} \|y - \overline{y}_j(\mathcal{C}_j)\|^2 = \sum_{y \in \mathcal{C}_1} \|y - \overline{y}_1(\mathcal{C}_1)\|^2 + \sum_{j=2}^{N-M} \sum_{y \in \mathcal{C}_j} \|y - \overline{y}_j(\mathcal{C}_j)\|^2 = \sum_{y \in \mathcal{C}_1} \|y - \overline{y}_1(\mathcal{C}_1)\|^2 = F(\mathcal{C}_1). \quad (8)$$

Отсюда следует, что алгоритм приближённого решения задачи MSSC–Case имеет следующий вид.

Шаг 1. По заданному множеству $\mathcal{Y}$ и числу M находим приближённое решение $\mathcal{B}^*$ задачи VS–2 с помощью алгоритма $\mathcal{A}$.

Шаг 2. Решением задачи MSSC-Case объявляем кластер $\mathcal{C}_1 = \mathcal{B}^*$ мощности M и совокупность $\{\mathcal{C}_2, \ldots, \mathcal{C}_{N-M+1}\} = \mathcal{Y} \backslash \mathcal{B}^*$ одноэлементных кластеров.

В силу (8), леммы 4 и теоремы 1 этот алгоритм гарантирует эффективное отыскание 2-приближённого решения задачи MSSC-Case.

Алгоритм $\mathcal{A}$ можно использовать для эффективного приближённого решения следующей NP-трудной [1] задачи.

Задача VS–1 (Vector Subset 1). *Дано*: множество $\mathcal{Y} = \{y_1, \ldots, y_N\}$ векторов из $\mathbb{R}^q$, натуральное число $M > 1$. *Найти:* подмножество $\mathcal{C} \subseteq \mathcal{Y}$ векторов такое, что целевая функция

$$Q(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \Big\| \sum_{y \in \mathcal{C}} y \Big\|^2 + \sum_{y \in \mathcal{Y} \backslash \mathcal{C}} \|y\|^2$$

максимальна при ограничении $|\mathcal{C}| = M$ на мощность подмножества $\mathcal{C}$.

Правомерность применения алгоритма $\mathcal{A}$ для решения этой задачи следует из связи между целевыми функциями задач VS–2 и

VS–1, установленной в [1]. Эта связь выражается формулой

$$F(\mathcal{C}) = \sum_{y\in\mathcal{Y}} \|y\|^2 - Q(\mathcal{C}),$$

в которой сумма является константой. Поэтому решение задачи VS–1 — максимизации $Q(\mathcal{C})$ — эквивалентно решению задачи VS–2 — минимизации $F(\mathcal{C})$. Следует, однако, заметить, что найденное с помощью алгоритма $\mathcal{A}$ решение задачи VS–1 не имеет каких-либо теоретических гарантий по точности решения.

Наконец, построенный алгоритм можно использовать для эффективного 2-приближённого решения следующей NP-трудной [1] задачи, тесно связанной с задачей VS–2.

Задача VS–3 (Vector Subset 3). *Дано*: множество $\mathcal{Y} = \{y_1, \dots, y_N\}$ векторов из $\mathbb{R}^q$, натуральное число $M > 1$. *Найти:* подмножество $\mathcal{C} \subseteq \mathcal{Y}$ векторов такое, что целевая функция

$$H(\mathcal{C}) = \sum_{y\in\mathcal{C}}\sum_{z\in\mathcal{C}} \|y - z\|^2 \tag{9}$$

минимальна при ограничении $|\mathcal{C}| = M$ на мощность искомого подмножества.

В самом деле, в силу известной [2] формулы

$$\sum_{y\in\mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 = \frac{1}{2|\mathcal{C}|}\sum_{y\in\mathcal{C}}\sum_{z\in\mathcal{C}} \|y - z\|^2$$

из (1) и (9) получаем $H(\mathcal{C}) = 2|\mathcal{C}| \cdot F(\mathcal{C})$. Поскольку $|\mathcal{C}| = M = \mathrm{const}$, для оптимального значения целевой функции задачи VS–3 справедливо равенство $H(\mathcal{C}^*) = 2M \cdot F(\mathcal{C}^*)$.

Возьмём в качестве приближённого решения этой задачи подмножество $\mathcal{B}^*$, найденное с помощью алгоритма $\mathcal{A}$. Для этого решения значение целевой функции задачи VS–3 равно $H(\mathcal{B}^*) = 2M \cdot F(\mathcal{B}^*)$. Поэтому в силу леммы 4 и теоремы 1 для точности решения задачи имеем оценку $H(\mathcal{B}^*)/H(\mathcal{C}^*) = F(\mathcal{B}^*)/F(\mathcal{C}^*) \leqslant 2$.

Заключение

Построен эффективный приближённый алгоритм с оценкой точности 2 для решения NP-трудной задачи, к которой сводится поиск в множестве векторов евклидова пространства подмножества, имеющего заданную мощность и включающего векторы, «близкие» между собой

по критерию минимума суммы квадратов расстояний. Установлено, что оценка точности 2 этого алгоритма достижима и неулучшаема. Показано, что предложенный алгоритм применим для отыскания 2-приближённых решений некоторых родственных задач анализа данных.

Обоснование алгоритмов другого типа (асимптотически точных, рандомизированных и др.) для решения рассмотренной и родственных задач является делом ближайшей перспективы. Интерес представляют алгоритмические решения задачи VS–1, поскольку в настоящее время для этой задачи какие-либо эффективные алгоритмы с оценками точности неизвестны.

ЛИТЕРАТУРА

1. **Кельманов А. В., Пяткин А. В.** NP-полнота некоторых задач выбора подмножества векторов // Дискрет. анализ и исслед. операций. — 2010. — Т. 17, №5. — С. 37–45.
2. **Edwards A. W. F., Cavalli–Sforza L. L.** A method for cluster analysis // Biometrics. — 1965. — Vol. 21. — P. 362–375.
3. **Wirth H.** Algorithms+data structures-programs. — New Jersey: Prentice Hall, 1976. — 366 p.

Александр Васильевич Кельманов,
e-mail: kelm@math.nsc.ru
Романченко Семён Михайлович,
e-mail: semenr@bk.ru

Статья поступила
26 июля 2010 г.

Переработанный вариант —
9 октября 2010 г.