ЗАДАЧА ВЫБОРА ПОДМНОЖЕСТВА ВЕКТОРОВ С МАКСИМАЛЬНОЙ СУММОЙ

Э. Х. Гимади

1. Введение. В докладе рассматриваются дискретные оптимизационные задачи, связанные с выбором из конечного семейства векторов в евклидовом пространстве R^k подмножества векторов с максимальной нормой суммы. В общем случае эти задачи NP-трудны, что стимулирует выделение таких подклассов этих задач, для которых оказывается возможным построение точных, либо приближенных алгоритмов полиномиальной или псевдополиномиальной временной сложности.

Полученные результаты могут быть использованы для решения задачи выбора фиксированного числа фрагментов в числовой последовательности квазипериодически повторяющегося фрагмента при заданном числе повторов. Подобная ситуация типична для таких приложений, как электронная разведка, радиолокация, телекоммуникация, геофизика, обработка речевых сигналов, медицинская и техническая диагностика и др. [2,4]

Далее под нормой будем понимать евклидову норму в k-мерном пространстве R^k, т. е. $\|x\| = \sqrt{x_1^2 + \ldots + x_k^2}$.

В качестве основных задач рассматриваем следующие две:

Задача 1: *Задано конечное семейство векторов $V = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ в евклидовом пространстве R^k и натуральное число $m < n$. Требуется найти подсемейство векторов из V мощности m, обладающее максимальной нормой суммы.*

Запишем задачу 1 в терминах целочисленного программирования:

$$\sum_{i=1}^{k}\Big(\sum_{j=1}^{n} v_{ij}x_j\Big)^2 \to \max; \tag{1}$$

$$\sum_{j=1}^{n} x_j = m; \tag{2}$$

$$x_j \in \{0,1\},\ 1 \le j \le n. \tag{3}$$

Задача 2: *Задано конечное семейство векторов $V = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ в евклидовом пространстве R^k и натуральные числа m и l, удовлетворяющие условию $lm < n$. Требуется выделить в V подсемейство векторов $X = \{\vec{v}_{a_1}, \vec{v}_{a_2}, \ldots, \vec{v}_{a_m}\}$, обладающее максимальной нормой суммы при условии $a_{i+1} - a_i \ge l$ для $i = 1, 2, \ldots, m-1$.*

В случае $k \le l$ вторая задача связана с целесообразным поиском сигналов в импульсной последовательности, зашумленной аддитивным шумом [2, 4].

2. Известные факты о сложности и алгоритмах решения задач 1 и 2.

Теорема 1. [1] *Задачи 1 и 2 NP-трудны.*

Теорема 2. [1] *Задача 1 решается с гарантированной относительной погрешностью, не превышающей величины $\frac{k-1}{8L^2}$ за время*

$$O\Big(nk^2(2L+1)^{k-1}\Big),$$

где L — параметр алгоритма.

Теорема 3. [1] *При фиксированной размерности k пространства R^k задача 1 решается асимптотически точно при выборе параметра $L = L(n)$, где $L(n)$ — произвольная неограниченно растущая функция от n.*

Тем самым для случая фиксированной размерности k пространства R^k установлено построение полиномиальной аппроксимационной схемы. Действительно, положим относительную погрешность равной $\varepsilon = \frac{k-1}{8L^2}$. Тогда $L = (\frac{k-1}{8\varepsilon})^{1/2}$ и для времени решения получим оценку

$$O\Big(nk^2\Big(\sqrt{\frac{k-1}{2\varepsilon}}+1\Big)^{k-1}\Big).$$

Теорема 4. [1] *Задача 2 решается с гарантированной относительной погрешностью, не превышающей $\frac{k-1}{8L^2}$, за время*

$$O\Big(nk(k+m)(2L+1)^{k-1}\Big).$$

Пусть b — максимальная по абсолютной величине координата векторов из семейства V.

Теорема 5. [1] *Если размерность k пространства R^k фиксирована, то задача 1 с целочисленными координатами входных векторов решается точно алгоритмом с параметром $L = 0,5kmb$ за псевдополиномиальное время $O(nk^2(kmb)^{k-1})$.*

3. Алгоритмы в духе динамического программирования для решения задачи 1 с целочисленными координатами векторов.

3.1. Случай неотрицательных целочисленных координат векторов.

Рассмотрим сначала случай векторов в положительном ортанте.

Обозначим через $\vec{B} \in Z_+^{k-1}$ вектор с компонентами $B_i = \sum_{r=1}^m v_{i,\sigma_r(i)},\ 1 \le i < k$, где $\sigma(i)$ — перестановка, упорядочивающая элементы $(v_{i1}, v_{i2}, \dots, v_{in})$ i-й строки матрицы (v_{ij}) по невозрастанию; $\mathcal{B} = \{\vec{\beta} \in Z_+^{k-1} \bigm| 0 \le \vec{\beta} \le \vec{B}\}$.

Теорема 6. *Пусть $f_{mn}(\vec{\beta})$ — максимум функции*

$$\Big\{\sum_{j=1}^n v_{kj}x_j \Bigm| \sum_{j=1}^n v_{ij}x_j = \beta_i,\ 1 \le i < k;\ \sum_{j=1}^n x_j = m;\ x_j \in \{0,1\},\ 1 \le j \le n\Big\}.$$

Тогда оптимум задачи 1 с целочисленными неотрицательными координатами векторов равен

$$S^* = \max\Big\{\sum_{i=1}^{k-1}\beta_i^2 + f_{mn}^2(\vec{\beta}) \Bigm| \vec{\beta} \in \mathcal{B}\Big\}.$$

Доказательство. Задачу 1 в форме (1)-(3) запишем в эквивалентном виде

$$\sum_{i=1}^{k-1}\beta_i^2 + \Big(\sum_{j=1}^n v_{kj}x_j\Big)^2 \to \max; \tag{4}$$

$$\sum_{j=1}^n v_{ij}x_j = \beta_i,\ 1 \le i < k; \tag{5}$$

$$\vec{\beta} \in \mathcal{B}; \tag{6}$$

$$\sum_{j=1}^{n} x_j = m; \tag{7}$$

$$x_j \in \{0,1\},\ 1 \le j \le n, \tag{8}$$

откуда непосредственно следует справедливость требуемого равенства.

Алгоритм $\widetilde{A}$ вычисления оптимумов $\{f_{mn}(\vec{\beta}) \;\Big|\; \vec{\beta} \in \mathcal{B}\}$.

Введем обозначение $< m, n;\ \vec{\beta} >$ для задачи отыскания оптимума $f_{mn}(\vec{\beta})$ и пусть $\{< \mu, j;\ \vec{\beta} > \Big|\ 1 \le \mu \le j \le n;\ 0 \le \vec{\beta} \le \vec{B}\}$ — семейство подзадач с соответствующими оптимумами $f_{\mu,j}(\vec{\beta})$.

Через $\widetilde{f}_{\mu,j}(\vec{\beta})$ обозначим оптимум в такой подзадаче $< \mu, j;\ \vec{\beta} >$, в которой выбирается μ векторов среди $\vec{v}_1, \ldots, \vec{v}_j$ с предписанным выбором последнего вектора $\vec{v}_j$.

Описание алгоритма $\widetilde{A}$.

Предварительный шаг:

a) $f_{1,j}(\vec{\beta}) := \infty;\ \widetilde{f}_{1,j}(\vec{\beta}) := \infty$ для всяких $j = 1, \ldots, n$ и $\vec{\beta} \in \mathcal{B}$.
b) $\widetilde{f}_{1,j}(\vec{v_j}) := v_{kj}$ для всякого $j = 1, \ldots, n$.
c) $f_{1,j}(\vec{\beta}) := \max\{\widetilde{f}_{1,j}(\vec{\beta});\ f_{1,j-1}(\vec{\beta})\}$ для всяких $j = 1, \ldots, n$ и $\vec{\beta} \in \mathcal{B}$.

Общий шаг $\mu = 2, \ldots, m$:

a) $\widetilde{f}_{\mu,j}(\vec{\beta}) := \infty$ для всяких $j = \mu, \ldots, n$ и $\vec{\beta} \in \mathcal{B}$.
b) Для всяких $\vec{\beta} \in \mathcal{B}$ положить

$$\widetilde{f}_{\mu,j}(\vec{\beta}) = \begin{cases} f_{\mu,\mu}(\vec{\beta}) & \text{при } j = \mu, \\ v_{k,j} + f_{\mu,j-1}(\vec{\beta} - \vec{v_j}), & \text{если } \mu < j \le n; \end{cases}$$

$$f_{\mu,j}(\vec{\beta}) = \begin{cases} \widetilde{f}_{\mu,\mu}(\vec{\beta}) & \text{при } j = \mu, \\ \max\{f_{\mu,j}(\vec{\beta});\ \widetilde{f}_{\mu,j-1}(\vec{\beta})\}, & \text{если } \mu < j \le n. \end{cases}$$

Из описания алгоритма $\widetilde{A}$ следует

Теорема 7. *Если размерность k пространства R^k фиксирована, то задача 1 с целочисленными неотрицательными координатами векторов решается точно за псевдополиномиальное время $O\Big(mn\prod_{i=1}^{k-1} B_i\Big)$.*

3.2. Случай входных векторов $\vec{v_1}, \ldots, \vec{v_n}$ с целочисленными координатами произвольного знака.

Модифицируем алгоритм $\widetilde{A}$ решения задачи 1 на случай входных векторов $\vec{v}_1, \ldots, \vec{v}_n$ с целочисленными координатами произвольного знака.

Пусть теперь $B_i = \sum_{r=1}^{m} (v_{i,\sigma_r(i)} - v_{i,\sigma_r(n-i+1)}),\ 1 \le i < k$, являются компонентами вектора $\vec{B} \in Z_+^{k-1}$; $\mathcal{B} = \{\vec{\beta} \in Z_+^{k-1} \mid 0 \le \vec{\beta} \le \vec{B}\}$; $b_i' = \min\{v_{ij} \ |1 \le j \le n\}\ 1 \le i \le k$.

Положив $v'_{ij} = v_{ij} - b'_i$, имеем матрицу (v'_{ij}) с неотрицательными целочисленными элементами. Применив алгоритм $\widetilde{A}$ к задаче 1 с модифицированной входной матрицей (v'_{ij}), получим оптимум исходной задачи со входом $(\vec{v}_1, \ldots, \vec{v}_n)$:

$$S^* = \max\Big\{\sum_{i=1}^{k-1}(\beta_i - mb'_i)^2 + \Big(f_{mn}(\vec{\beta}) - mb'_k\Big)^2 \ \Big| \ \vec{\beta} \in \mathcal{B}\Big\}.$$

Заметим, что утверждение теоремы 6 в случае координат произвольного знака остается в силе. Однако есть отличие в оценке временной сложности, если оценить компоненты вектора $\vec{B}$ в терминах величины b (максимального абсолютного значения среди координат входных векторов в семействе V):

Теорема 8. *Если размерность k пространства R^k фиксирована, то задача 1 с целочисленными координатами векторов решается точно за псевдополиномиальное время $O\Big(mn(mb)^{k-1}\Big)$ (в случае неотрицательных координат) и $O\Big(mn(2mb)^{k-1}\Big)$ (в случае координат произвольного знака).*

Заключительные замечания.

1) Аналогичные алгоритмические построения в духе динамического рограммирования применимы также для решения задачи 2.

2) Остается открытым вопрос о сложностном статусе основных задач 1 и 2 при фиксированной размерности пространства R^k.

3) Представленные в докладе результаты получены в ходе совместной работы по грантам РФФИ и ИНТАС (проект 04-77-7173) с моими коллегами — А.Е. Бабуриным, Н.И. Глебовым и А.В. Пяткиным [1].

4) Наше внимание к рассмотрению данного класса задач дискретной оптимизации было инициировано А.А. Кельмановым [4].

Работа поддержана грантами РФФИ (коды проектов 05-01-00395, 07-07-00022, 07-07-00168).

ЛИТЕРАТУРА

1. А. Е. Бабурин, Э. Х. Гимади, Н. И. Глебов, А. В. Пяткин. Задача отыскания подмножества векторов с максимальным суммарным весом // Дискрет. анализ и исслед. операций, Серия 2, 2007, Т. 14, N 1.

2. Э.Х. Гимади, А.В. Кельманов, М.А. Кельманова, С.А. Хамидуллин. Апостериорное обнаружение в числовой последовательности квазипериодически повторяющегося фрагмента при заданном числе повторов // Сиб. журн. индустр. математики. 2006. Т. IX, N(25). С. 55–74.

3. М. Гэри, Д. Джонсон. Вычислительные машины и труднорешаемые задачи М.: Мир. 1982.

4. А.В. Кельманов, М.А. Кельманова, С.А. Хамидуллин. Апостериорное обнаружение одинаковых подпоследовательностей-фрагментов в квазипериодической последовательности // Сиб. журн. индустр. математики. 2002. Т. V, N 2(10). С. 94-108.

Гимади Эдуард Хайрутдинович, Институт математики им. С.Л. Соболева СО РАН, пр. академика Коптюга, 4, Новосибирск, 630090, Россия, тел. (8-383-3) 33-21-89, факс (8-383-3) 33-25-98, E-mail:gimadi@math.nsc.ru