ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь—февраль 2013. Том 20, № 1. С. 93–99

УДК 519.176

ЗАДАЧА О МИНИМАЛЬНОМ ШАРЕ, ОХВАТЫВАЮЩЕМ k ТОЧЕК *)

В. В. Шенмайер

Аннотация. Рассматривается задача поиска шара минимального радиуса, охватывающего не менее k точек из заданного конечного множества в евклидовом пространстве. В случае фиксированной размерности пространства задача полиномиально разрешима, но в общем случае сложностной статус задачи до настоящего времени не был установлен. Доказано, что задача NP-трудна в сильном смысле, а также получена полиномиальная аппроксимационная схема (PTAS), позволяющая решать задачу с произвольной относительной погрешностью ε за время $O(n^{1/\varepsilon^2+1}d)$, где n — мощность исходного множества, d — размерность пространства.

Ключевые слова: минимальный охватывающий шар, кластерный анализ, аппроксимационная схема, приближённый алгоритм.

Введение

Рассмотрим задачу о минимальном шаре, охватывающем k точек.

Задача Sk-EB (Smallest k-Enclosing Ball).
Дано: конечное множество X в евклидовом пространстве $\mathbb{R}^d$ и целое число $k \in [1, n]$, где $n = |X|$.
Найти: шар минимального радиуса, охватывающий не менее k точек из X.

Задача имеет множество содержательных интерпретаций, что обусловлено простотой и наглядностью формулировки. Одна из них лежит в области военного дела: заданы координаты n целей, требуется поразить k из них залпом минимального заряда.

Первая математическая постановка одного из частных случаев задачи появилась ещё в середине XIX века [8]. В случае фиксированной

*) Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проекты 09–01–00032 и 10–07–00195), целевой программы № 2 Президиума РАН (проект № 227), а также целевой программы СО РАН (интеграционный проект № 44).

размерности пространства, в частности, на плоскости (наиболее исследованный случай), задача полиномиально разрешима [2]. При этом трудоёмкость наилучших известных алгоритмов, решающих данную задачу, зависит линейно от мощности исходного множества и экспоненциально от размерности пространства [4, 6]. В общем случае, когда размерность пространства не фиксирована (является частью входа), сложностной статус задачи до настоящего времени не установлен.

В данной статье получены следующие результаты. Доказано, что задача NP-трудна в сильном смысле. Установлено, что при условии P$\neq$NP для неё не существует полностью полиномиальной аппроксимационной схемы (FPTAS). Получена полиномиальная аппроксимационная схема (PTAS), позволяющая решать данную задачу с произвольной относительной погрешностью ε за время $O(n^{1/\varepsilon^2+1}d)$.

1. Анализ сложности

Сформулируем частный случай рассматриваемой оптимизационной задачи в форме задачи верификации свойств на множестве булевых векторов.

Дано: конечное множество векторов $X \subseteq \{0,1\}^d$, целое $k \in [1,n]$ и положительное R.
Вопрос: существует ли шар радиуса R в евклидовой метрике, охватывающий не менее k точек из множества X?

Воспользовавшись идеями из [1], сведём к данному частному случаю следующую NP-полную в сильном смысле задачу [7].

Задача о k-клике в однородном графе.
Дано: однородный граф G и целое число k.
Вопрос: существует ли в графе G такое подмножество вершин мощности k, что любые две вершины из этого подмножества связаны ребром?

Теорема 1. *Задача о k-клике в однородном графе полиномиально сводится к частному случаю задачи о минимальном шаре, охватывающем k точек.*

Схема сведéния. Пусть G — однородный граф степени s на n вершинах с m рёбрами. Построим следующий пример задачи о существовании шара. Каждой вершине графа G сопоставим m-мерный вектор, в котором i-я координата равна 1, если i-е ребро инцидентно этой вершине, и 0 в противном случае. Пусть множество X состоит из определённых таким образом векторов.

Лемма 1. *Пусть K — произвольное k-элементное множество в $\mathbb{R}^d$, $\overline{c}(K) = \sum\limits_{x\in K} x$, k — среднее множества K и $f(y,K) = \sum\limits_{x\in K} \|x-y\|^2$, где $y \in \mathbb{R}^d$. Тогда*

$$f(y,K) = f(\overline{c}(K),K) + k\,\|y-\overline{c}(K)\|^2. \tag{1}$$

Доказательство. Действительно,

$$f(y,K) = \sum_{x\in K} \|x-\overline{c}(K)\|^2 - \sum_{x\in K} 2\,\langle x-\overline{c}(K), y-\overline{c}(K)\rangle + \sum_{x\in K} \|y-\overline{c}(K)\|^2,$$

где $\langle .\,,.\rangle$ — скалярное произведение. Первое слагаемое в этом выражении равно $f(\overline{c}(K),K)$, последнее — $k\,\|y-\overline{c}(K)\|^2$, а второе — нулю в силу того, что $\sum\limits_{x\in K} x = k\,\overline{c}(K)$. Лемма 1 доказана.

Лемма 2. *Пусть $g(K) = \sum\limits_{x\in K}\sum\limits_{y\in K} \|x-y\|^2$. Тогда*

$$g(K) = 2k\,f(\overline{c}(K),K). \tag{2}$$

Утверждение леммы 2 следует из равенства (1).

Лемма 3. *Пусть R — радиус минимального шара, охватывающего k точек из множества X, и $A = (1-1/k)(s-1)$. Тогда $R^2 \leqslant A$, если в графе G существует k-клика, и $R^2 \geqslant A + 1/k^2$ в противном случае.*

Доказательство. Допустим граф G содержит клику на k вершинах. Рассмотрим в качестве центра искомого шара точку $\overline{c}(K)$. Из однородности графа G следует, что $\|x-y\|^2 = 2s$ для любой пары несовпадающих точек x, y из множества X, если соответствующие этим точкам вершины в графе G несмежны, и $\|x-y\|^2 = 2s-2$, если эти вершины смежны. Отсюда согласно формуле (2) имеем

$$f(\overline{c}(K),K) = g(K)/2k = (k^2-k)(2s-2)/2k = k\,A,$$

где K — множество векторов, соответствующих вершинам клики. В силу симметрии расстояния от точки $\overline{c}(K)$ до всех точек из K равны. Следовательно, квадраты этих расстояний равны величине $f(\overline{c}(K),K)/k = A$. Таким образом, шар радиуса $\sqrt{A}$ с центром в точке $\overline{c}(K)$ охватывает все k точек множества K.

Допустим в графе G не существует k-клики. Предположим, что c — центр минимального шара, охватывающего k точек, и K — охватываемое им k-элементное подмножество множества X. Тогда R равно расстоянию

от точки c до самой дальней от неё точки из K. В силу того, что среднее значение квадратов расстояний не превосходит максимального значения, имеем $R^2 \geqslant f(c, K)/k$. С другой стороны, $f(c, K) \geqslant f(\overline{c}(K), K)$ согласно равенству (1). Следовательно, $R^2 \geqslant f(\overline{c}(K), K)/k = g(K)/2k^2$. Но каждое из $k^2 - k$ слагаемых в определении величины $g(K)$, соответствующих парам несовпадающих точек, равно либо $2s$, либо $2s - 2$. Причём в силу предположения о несуществовании k-клики по крайней мере одно из этих слагаемых равно $2s$. Таким образом, $g(K) \geqslant (k^2 - k)(2s - 2) + 2$, откуда следует, что $R^2 \geqslant (1 - 1/k)(s - 1) + 1/k^2 = A + 1/k^2$. Лемма 3 доказана.

Доказательство теоремы 1. Из леммы 3 следует, что k-клика в графе G существует тогда и только тогда, когда существует шар радиуса $\sqrt{A}$, охватывающий k точек из X. Теорема 1 доказана.

Следствие. *Задача о минимальном шаре, охватывающем k точек,* NP-*трудна в сильном смысле.*

В случае целочисленной полиномиально-ограниченной целевой функции из NP-трудности в сильном смысле следует, что не существует полностью полиномиальной аппроксимационной схемы (FPTAS) [5, 9]. Однако радиус шара, охватывающего k-элементное подмножество, не является целочисленной функцией, поэтому в рассматриваемом случае данный факт требует доказательства.

Теорема 2. *При условии* P≠NP *для задачи о минимальном шаре, охватывающем k точек, не существует полностью полиномиальной аппроксимационной схемы* (FPTAS).

Доказательство. Согласно лемме 3 радиус минимального шара, охватывающего k точек из множества X, в случае существования k-клики в графе G не превосходит величины $\sqrt{A}$, а в случае её несуществования оценивается снизу величиной

$$\sqrt{A + 1/k^2} > \sqrt{A}(1 + 1/3Ak^2) > \sqrt{A}(1 + 1/3n^3).$$

Следовательно, при условии P≠NP не существует приближённого полиномиального алгоритма с относительной погрешностью $1/3n^3$. С другой стороны, полностью полиномиальная аппроксимационная схема позволяет за полиномиальное время получать решения с любой относительной погрешностью вида $1/p(n)$, где $p(n)$ — произвольный полином от длины входа задачи. Отсюда следует, что такой схемы не существует. Теорема 2 доказана.

2. Аппроксимационная схема

Предлагаемая полиномиальная аппроксимационная схема (PTAS) для задачи о минимальном шаре, охватывающем k точек, опирается на простой алгоритм типа градиентного спуска [3] для задачи о минимальном охватывающем шаре.

Задача SEB (Smallest Enclosing Ball).
Дано: конечное множество K в евклидовом пространстве $\mathbb{R}^d$.
Найти: шар минимального радиуса, охватывающий все точки из K.

Заметим, что эта задача является частным случаем задачи Sk-EB (в данном случае $k = n$). В [3] предложен следующий итерационный приближённый алгоритм для её решения.

Алгоритм.

Пусть i — произвольное целое положительное число, параметр алгоритма.

Шаг 1. Выбираем произвольную точку $c_1 \in K$.

Шаг j, $j = 2, 3, \ldots, i$. Находим точку $p_j \in K$, наиболее удалённую от точки c_{j-1}, и полагаем $c_j = c_{j-1} + (p_j - c_{j-1})/i$.

Выход алгоритма: шар с центром c_i и радиусом

$$R(c_i, K) = \max_{x \in K} \|x - c_i\|.$$

Отметим, что точки $p_1, p_2, \ldots, p_i$, где $p_1 = c_1$, могут повторяться. При этом точка c_i равна среднему точек $p_1, \ldots, p_i$, последовательно выбранных на шагах алгоритма с 1-го по i-й, $c_i = \sum\limits_{j=1}^{i} p_j/i$.

Утверждение 1 [3]. *Пусть c^* и R^* — центр и радиус минимального шара, охватывающего множество K. Тогда $\|c_i - c^*\| \leqslant R^*/\sqrt{i}$.*

Из утверждения 1 и неравенства треугольника следует, что шар с центром c_i и радиусом $R(c_i, K)$ является $(1 + 1/\sqrt{i}\,)$-приближённым решением задачи SEB. Таким образом, описанный выше алгоритм является полностью приближённой аппроксимационной схемой (FPTAS), позволяющей решать задачу о минимальном охватывающем шаре с произвольной относительной погрешностью $\varepsilon > 0$ за время $O(nd/\varepsilon^2)$.

Построим алгоритм для исходной задачи о минимальном шаре, охватывающем k точек (Sk-EB). Пусть c^* и R^* — центр и радиус шара, являющегося оптимальным решением этой задачи. Очевидно, что этот

шар является также и оптимальным решением задачи SEB на k-элементном множестве K^* охватываемых им точек. Следовательно, в силу утверждения 1 среднее некоторых (возможно повторяющихся) точек $p_1, \ldots, p_i \in K^*$ лежит на расстоянии $R^*/\sqrt{i}$ от точки c^*. О точках $p_1, \ldots, p_i$, как и о множестве K^*, ничего изначально не известно, кроме того, что они принадлежат множеству X. Идея алгоритма заключается в переборе всех последовательностей точек множества X длины i, чтобы в числе всех вариантов рассмотреть и нужную последовательность $p_1, \ldots, p_i$ и соответственно её среднее, близкое к центру оптимального шара.

Алгоритм решения задачи Sk-EB.

Пусть $t^i : \{1, \ldots, n^i\} \to X^i$ — перечисление всех последовательностей точек множества X длины i, где i — произвольное целое положительное число, параметр алгоритма.

Шаг s, $s = 1, \ldots, n^i$. Рассмотрим последовательность $t^i(s)$. Пусть $t^i(s) = p_1, \ldots, p_i$. Положим $c^i(s) = \sum\limits_{j=1}^{i} p_j/i$. Находим множество $K^i(s)$, состоящее из k ближайших к точке $c^i(s)$ точек X, и определим радиус $R(c^i(s), K^i(s)) = \max\limits_{x \in K^i(s)} \|x - c^i(s)\|$.

Выход алгоритма: шар с центром c^i и радиусом R^i, соответствующий наименьшему из радиусов $R(c^i(s), K^i(s))$, $s = 1, \ldots, n^i$.

Теорема 3. *Пусть c^* и R^* — центр и радиус минимального шара, охватывающего k точек множества X. Тогда $R^i/R^* \leqslant 1 + 1/\sqrt{i}$.*

Доказательство. Как отмечено выше, оптимальное решение задачи Sk-EB является также и оптимальным решением задачи SEB на k-элементном множестве K^* охватываемых им точек. Пусть точки $p_1, \ldots, p_i$ выбраны на шагах $1, \ldots, i$ алгоритма решения задачи SEB на множестве K^*, $c_i = \sum\limits_{j=1}^{i} p_j i$ и $R(c_i, K^*) = \max\limits_{x \in K^*} \|x - c_i\|$. Тогда

$$R(c_i, K^*)/R^* \leqslant 1 + 1/\sqrt{i}$$

согласно утверждению 1 и неравенству треугольника.

С другой стороны, существует последовательность $t^i(s)$, определяемая алгоритмом для задачи Sk-EB, равная $p_1, \ldots, p_i$. Тогда $c^i(s) = c_i$ и $R(c^i(s), K^i(s)) = R(c_i, K^i(s)) \leqslant R(c_i, K^*)$, поскольку множество $K^i(s)$ состоит из точек, ближайших к точке c_i. Следовательно, $R^i \leqslant R(c_i, K^*)$, откуда и получаем, что $R^i/R^* \leqslant 1 + 1/\sqrt{i}$. Теорема 3 доказана.

Оценим трудоёмкость алгоритма. Поскольку выбор k ближайших к точке $c^i(s)$ точек множества X занимает не более n действий (например, с помощью предложенного в [10] алгоритма поиска k-го наименьшего числа в массиве из n чисел), а все арифметические операции с векторами линейно зависят от размерности пространства, трудоёмкость данного алгоритма оценивается величиной $O(n^{i+1}d)$.

Отметим, что при $i = 1/\varepsilon^2$, где $\varepsilon > 0$ — произвольная заданная относительная погрешность, трудоёмкость алгоритма оценивается величиной $O(n^{1/\varepsilon^2+1}d)$. Таким образом, описанный алгоритм является полиномиальной аппроксимационной схемой (PTAS) для задачи о минимальном шаре, охватывающем k точек.

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. **Aggarwal A., Imai H., Katoh N., Suri S.** Finding k points with minimum diameter and related problems // J. Algorithms. — 1991. — Vol. 12. — P. 38–56.
3. **Badoiu M., Clarkson K. L.** Smaller core-sets for balls // Proc. 14th ACM-SIAM Symposium on Discrete Alg. (Baltimore, January 12–14, 2003). — Philadelphia: SIAM, 2003. — P. 801–802.
4. **Eppstein D., Erickson J.** Iterated nearest neighbors and finding minimal polytopes // Discrete Comput. Geom. — 1994. — Vol. 11. — P. 321–350.
5. **Garey M. R., Johnson D. S.** “Strong” NP-completeness results: motivation, examples, and implications // J. ACM. — 1978. — Vol. 25, N 3. — P. 499—508.
6. **Har-Peled S., Mazumdar S.** Fast algorithms for computing the smallest k-enclosing disc // Algorithmica. — 2005. — Vol. 41, N 3. — P. 147–157.
7. **Papadimitriou C. H.** Computational complexity. — New York: Addison-Wesley, 1994. — 523 p.
8. **Sylvester J. J.** A question in the geometry of situation // Quart. J. Math. — 1857. — Vol. 1. — P. 79.
9. **Vazirani V.** Approximation algorithms. — Berlin: Springer-Verl., 2001. — 71 p.
10. **Wirth H.** Algorithms + data structures = programs. — New Jersey: Prentice Hall, 1976. — 366 p.

Шенмайер Владимир Владимирович,
e-mail: shenmaier@mail.ru

Статья поступила
26 июля 2012 г.