ПРИБЛИЖЁННЫЙ ПОЛИНОМИАЛЬНЫЙ АЛГОРИТМ ДЛЯ ОДНОЙ ЗАДАЧИ БИКЛАСТЕРИЗАЦИИ ПОСЛЕДОВАТЕЛЬНОСТИ[1]

А.В. Кельманов, С.А. Хамидуллин

Институт математики им. С.Л. Соболева Сибирского отделения РАН,
Новосибирский государственный университет, Новосибирск,
e-mail: kelm@math.nsc.ru, kham@math.nsc.ru

Рассматривается следующая NP-трудная в сильном смысле [1]

Задача. *Дано: последовательность* $\mathcal{Y} = (y_1, \ldots, y_N)$ *векторов из* $\mathbb{R}^q$, *натуральные числа* $T_{\min}$ *и* $T_{\max}$. *Найти: подмножество* $\mathcal{M} = \{n_1, \ldots, n_M\} \subseteq \mathcal{N} = \{1, \ldots, N\}$ *номеров элементов последовательности* $\mathcal{Y}$ *такое, что*

$$\sum_{j \in \mathcal{M}} \|y_j - \overline{y}(\mathcal{M})\|^2 + \sum_{i \in \mathcal{N} \setminus \mathcal{M}} \|y_i\|^2 \to \min,$$

где $\overline{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$, *при ограничениях*

$$1 \le T_{\min} \le n_m - n_{m-1} \le T_{\max} \le N, \; m = 2, \ldots, M,$$

на элементы набора $\mathcal{M}$.

Задача заключается в разбиении конечной последовательности векторов евклидова пространства на два кластера по критерию минимума суммы квадратов расстояний от элементов кластеров до их центров. Центр первого кластера является оптимизируемой величиной и определяется как среднее значение по всем векторам, образующим этот кластер. Центр второго кластера фиксирован в начале координат. При этом разбиение подчинено условию: разность между номерами последующего и предыдущего векторов, входящих в первый кластер, ограничена сверху и снизу заданными константами. Эта задача индуцируется, в частности, актуальными проблемами помехоустойчивого анализа данных (см. [1], [2] и цитированные там работы).

В работе предложен 2-приближенный полиномиальный алгоритм решения задачи. Алгоритм реализует схему динамического программирования и выполняется за время $\mathcal{O}(N^2(T_{\max} - T_{\min} + q))$. Поскольку $T_{\max} - T_{\min} < N$, алгоритм полиномиален и его трудоемкость можно оценить как $\mathcal{O}(N^2(N+q))$ в общем случае, и как $\mathcal{O}(N^2 q)$ в частном случае, когда $T_{\max} = T_{\min}$.

ЛИТЕРАТУРА

1. Кельманов А. В., Пяткин А. В. *О сложности некоторых задач кластерного анализа векторных последовательностей.* — Дискрет. анализ и исслед. операций. — 2013, Т. 20, № 2, с. 47–57.

2. Кельманов А. В., Пяткин А. В. *О сложности некоторых задач поиска подмножеств векторов и кластерного анализа.* — Журн. вычисл. математики и мат. физики. — 2009, Т. 49, № 11, с. 2059–2067.

[1]Работа выполнена при поддержке РФФИ (проекты № 12-01-00090 и № 13-07-00070)