An approximation polynomial algorithm for a problem of a sequence bi-partitioning[1]

Alexander Kel'manov, Sergey Khamidullin

Sobolev Institute of Mathematics, Novosibirsk State University, Novosibirsk,
e-mail: kelm@math.nsc.ru, kham@math.nsc.ru

We consider following strongly NP-hard [1]

Problem. *Given a sequence $\mathcal{Y} = (y_1, \dots, y_N)$ of vectors from $\mathbb{R}^q$, and some positive integer numbers $T_{\min}$ and $T_{\max}$. Find a subset $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ such that*

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

where $\overline{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|}\sum_{i\in\mathcal{M}} y_i$, under constraints

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

on the elements of $\mathcal{M}$.

The problem is to find a partition of a finite Euclidean vectors sequence into two clusters minimizing the sum of squared distances from the clusters elements to their centers. The center of the first cluster is defined as the mean values of all vectors in a cluster. The center of the second cluster is given in advance and is equal to 0. Additionally, the partition has to satisfy the following condition: for all vectors that are in the first cluster the difference between the indices of two consequent vectors from this cluster is bounded from below and above by some constants.

In this work we present a 2-approximation efficient algorithm for the problem. This algorithm implements a dynamic programming scheme and runs in time $O(N^2(T_{\max} - T_{\min} + q))$, where $T_{\max} - T_{\min} < N$. Therefore, algorithm is polynomial and it finds the solution in $O(N^2(N + q))$ time in the general case, and in $O(qN^2)$ time in the special case when $T_{\min} = T_{\max}$.

REFERENCES

1. A.V. Kel'manov, A.V. Pyatkin. *On Complexity of Some Problems of Cluster Analysis of Vector Sequences.* — J. Appl. Indust. Math. — 2013, Vol. 7, №3, p. 363–369.

2. A.V. Kel'manov, A.V. Pyatkin. *Complexity of Certain Problems of Searching for Subsets of Vectors and Cluster Analysis.* — Comput. Math. Math. Phys. — 2009, Vol. 49, №11, p. 1966–1971.

[1]The authors were supported by the Russian Foundation for Basic Research (projects no. 12-01-00090, no. 13-07-00070)