AN EXACT PSEUDOPOLYNOMIAL ALGORITHM FOR A TWO-CLUSTER PARTITIONING PROBLEM[1]

Alexander Kel'manov, Vladimir Khandeev

Sobolev Institute of Mathematics, Novosibirsk State University, Novosibirsk,
e-mail: kelm@math.nsc.ru, vladimir.handeev@gmail.com

In this paper we consider the following strongly NP-hard [1]

Problem. *Given a set* $\mathcal{Y} = \{y_1, \ldots, y_N\}$ *of vectors from* $\mathbb{R}^q$ *and a natural integer number* M*. Find a partition of* $\mathcal{Y}$ *into clusters* $\mathcal{C}$ *and* $\mathcal{Y} \setminus \mathcal{C}$ *such that*

$$\sum_{y \in \mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 + \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y\|^2 \to \min,$$

where $\overline{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum\limits_{y \in \mathcal{C}} y$ *is the center of cluster* $\mathcal{C}$*, under constrain* $|\mathcal{C}| = M$*.*

In [2], a 2-approximation algorithm for the problem is proposed. The running time of the algorithm is $O(qN^2)$. In [3], a polynomial-time approximation scheme with a $O(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$-time complexity, where ε is an arbitrary relative error, is substantiated. In [4], a randomized algorithm for the problem is presented. The running time of the algorithm for the fixed failure probability, relative error of the solution and for the certain value of parameter k is $\mathcal{O}(2^k q(k+N))$. The algorithm has also been proven to be asymptotically exact and to have $\mathcal{O}(qN^2)$-time complexity for the special values of the parameters.

In this work we present a pseudopolynomial algorithm which finds an optimal solution in the case of integer values of the components of the vectors in the input set and fixed space dimension. The running time of the algorithm is $\mathcal{O}(qN(2MD+1)^q)$, where D is the maximum absolute coordinate value of the vectors in the input set.

REFERENCES

1. A.V. Kel'manov, A.V. Pyatkin *Complexity of Certain Problems of Searching for Subsets of Vectors and Cluster Analysis.* — Computational Mathematics and Mathematical Physics. — 2009, vol.49, №11, p. 1966-1971.
2. A.V. Dolgushev, A.V. Kel'manov *An Approximation Algorithm for Solving a Problem of Cluster Analysis.* — Journal of Applied and Industrial Mathematics. — 2011, vol.5, №4, p. 551-558.
3. A.V. Dolgushev, A.V. Kel'manov, V.V. Shenmaier *PTAS for one problem of cluster analysis.* Intelligent Information Processing: Proceedings of the 9th International Conference. Montenegro, Budva. 2012, p. 242-244. (In Russian).
4. Alexander Kel'manov, Vladimir Khandeev *A randomized algorithm for a clustering problem.* Proceedings of IV International Conference «Optimization and applications» (OPTIMA-2013), Petrovac, Montenegro, September 22–28, 2013, p. 86.

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