An Approximation Scheme for a Weighted 2-Clustering Problem

Alexander Kel'manov[1,2], Anna Motkova[2], and Vladimir Shenmaier[1]

[1] Sobolev Institute of Mathematics,
4 Koptyug Ave., 630090 Novosibirsk, Russia
[2] Novosibirsk State University,
2 Pirogova St., 630090 Novosibirsk, Russia
kelm@math.nsc.ru,{anitamo,shenmaier}@mail.ru

We consider the following clustering problem:

Given an N-element set $\mathcal{Y}$ of points from $\mathbb{R}^q$, a positive integer $M \le N$, and real numbers (weights) $\omega_1 > 0$ and $\omega_2 \ge 0$. *Find* a partition of $\mathcal{Y}$ into two clusters $\mathcal{C}$ and $\mathcal{Y}\setminus\mathcal{C}$ minimizing the value of

$$\omega_1 \sum_{y\in\mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 + \omega_2 \sum_{y\in\mathcal{Y}\setminus\mathcal{C}} \|y\|^2,$$

where $|\mathcal{C}| = M$ and $\overline{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|}\sum_{y\in\mathcal{C}} y$ is the centroid of $\mathcal{C}$.

We present an algorithm that allows to find a $(1+\varepsilon)$-approximate solution of the problem in $\mathcal{O}\Big(\sqrt{q}N^2\big(\frac{\pi e}{2}\big)^{q/2}\big(\frac{1}{\sqrt{\varepsilon}}+2\big)^q\Big)$ time for any $\varepsilon \in (0,1)$. The algorithm implements a FPTAS in the case of fixed space dimension and remains polynomial for instances of dimension $\mathcal{O}(\log n)$.

Earlier, in [1–3], for the strongly NP-hard cases when (1) $\omega_1 = 1$ and $\omega_2 = 0$, (2) $\omega_1 = \omega_2 = 1$ and (3) $\omega_1 = |\mathcal{C}|$ and $\omega_2 = N - |\mathcal{C}|$, the approximation schemes with running time $\mathcal{O}\Big(qN^2\big(\sqrt{\frac{2q}{\varepsilon}}+1\big)^q\Big)$ were proposed.

Acknowledgments. This work was supported by the Russian Science Foundation (project 16-11-10041).

References

1. Kel'manov, A. V., Romanchenko, S. M.: An FPTAS for a Vector Subset Search Problem. J. Appl. Indust. Math. 8(3), 329–336 (2014)
2. Kel'manov, A. V., Khandeev, V. I.: Fully Polynomial-Time Approximation Scheme for a Special Case of a Quadratic Euclidean 2-Clustering Problem. Comput. Math. Math. Phys. 56(2), 334–341 (2016)
3. Kel'manov, A. V., Motkova, A. V.: A Fully Polynomial-Time Approximation Scheme for a Special Case of a Balanced 2-Clustering Problem. LNCS. 9869, 182–192 (2016)