

MSC: 68W25, 68Q25

DOI: 10.21538/0134-4889-2018-24-4-189-198

## ON THE COMPLEXITY OF SOME MAX–MIN CLUSTERING PROBLEMS

A. V. Kel'manov, A. V. Pyatkin, V. I. Khandeev

Two similar problems of searching for a family of disjoint subsets (clusters) in a finite set of points in Euclidean space are considered. In these problems the size of the smallest cluster should be maximized so that in each cluster the intracluster quadratic variation of the points with respect to its center would not exceed a given (constant) fraction of the total quadratic variation of the points of the input set with respect to its centroid. In the first problem the centers of intracluster variations are arbitrary points given at the input. In the second problem the centers of the intracluster variation are unknown (to be found) but they must lie in the input set. It is proved that both problems are NP-hard even on the real line both in the general case when the number of the clusters is a part of the input and in the parametric case when the number of the clusters is fixed.

Keywords: Euclidean space, clustering, max–min problem, quadratic variation, NP-hardness.

### REFERENCES

1. James G., Witten D., Hastie T., Tibshirani R. *An Introduction to Statistical Learning*. N Y: Springer Science+Business Media, LLC, 2013, 426 p. ISBN: 978-1461471370.
2. Aggarwal C.C. *Data Mining: The Textbook*. N Y: Springer International Publishing, 2015, 734 p. doi: 10.1007/978-3-319-14142-8.
3. Bishop C.M. *Pattern Recognition and Machine Learning*. New York: Springer Science+Business Media, LLC, 2006, 738 p. ISBN: 978-0-387-31073-2.
4. Shirkorshidi A.S., Aghabozorgi S, Wah T.Y., and Herawan T. Big data clustering: A review. In: Murgante B. et al. (eds) *Computational Science and Its Applications - ICCSA 2014. Lecture Notes in Computer Science*, 2014, vol. 8583, pp. 707–720. doi: 10.1007/978-3-319-09156-3\_49.
5. Alessio Farcomeni, Greco Luca. *Robust methods for data reduction*. Boca Raton: CRC Press, 2015, 297 p. ISBN: 9781466590625.
6. Osborne, Jason W. *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*, Los Angeles: SAGE Publication, Inc., 2013, 296 p. ISBN: 9781412988018.
7. Aloise D., Deshpande A., Hansen P., Popat P. NP-hardness of Euclidean sum-of-squares clustering. *Machine Learning*, 2009, vol. 75, no. 2, pp. 245–248. doi: 10.1007/s10994-009-5103-0.
8. Papadimitriou C.H. Worst-case and probabilistic analysis of a geometric location problem. *SIAM J. Comput.*, 1981, vol. 10, no. 3, pp. 542–557. doi: 10.1137/0210040.
9. Masuyama S., Ibaraki T., Hasegawa T. The computational complexity of the m-center problems in the plane. *Trans. IECE Jpn.*, 1981, vol. 64, no. 2, pp. 57–64.
10. Hatami B., Zarrabi-Zade H. A streaming algorithm for 2-center with outliers in high dimensions. *Computational Geometry*, 2017, vol. 60, pp. 26–36. doi: 10.1016/j.comgeo.2016.07.002.
11. Kel'manov A.V., Pyatkin A.V. NP-hardness of some Euclidean problems of partitioning a finite set of points. *Comput. Math. Math. Phys.*, 2018, vol. 58, no. 5, pp. 822–826. doi: 10.1134/S0965542518050123.
12. Garey M.R., Johnson D.S. *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco: Freeman, 1979, 340 p. ISBN: 978-0716710455.

Received July 30, 2018

Revised October 08, 2018

Accepted November 12, 2018

**Funding Agency:** This work was supported by the Russian Science Foundation (project no. 16-11-10041).

*Alexander Vasil'evich Kel'manov*, Dr. Phys.-Math. Sci., Sobolev Institute of Mathematics; Novosibirsk State University, Novosibirsk, 630090 Russia, e-mail: kelm@math.nsc.ru .

*Artem Valer'evich Pyatkin*, Dr. Phys.-Math. Sci., Sobolev Institute of Mathematics; Novosibirsk State University, Novosibirsk, 630090 Russia, e-mail: artem@math.nsc.ru .

*Vladimir Il'ich Khandeev*, Cand. Sci. (Phys.-Math.), Sobolev Institute of Mathematics; Novosibirsk State University, Novosibirsk, 630090 Russia, e-mail: khandeev@math.nsc.ru .

Cite this article as:

A. V. Kel'manov, A. V. Pyatkin, V. I. Khandeev. On the complexity of some max–min clustering problems, *Trudy Inst. Mat. Mekh. UrO RAN*, 2018, vol. 24, no. 4, pp. 189–198 .