TRUDY INSTITUTA MATEMATIKI I MEKHANIKI UrO RAN

Vol. 23 No. 4 **2017**

MSC: 90C27, 90C05, 62H30

DOI: 10.21538/0134-4889-2017-23-4-301-310

ATTAINABLE BEST GUARANTEE FOR THE ACCURACY OF K-MEDIANS CLUSTERING IN $[0,1]$

M. Yu. Khachai, D. M. Khachai, V. S. Pankratov

The scalar k-medians clustering problem is considered in the context of a two-player zero-sum game. The set of strategies of the first player coincides with a family of fixed-length samples from the interval $[0,1]$. The strategies of the second player are all possible partitions of an arbitrary sample of a given length into a given number of clusters. The quality of the clustering is evaluated by the payoff function equal to the sum of deviations of the elements from the centers of clusters nearest to them. It is easy to see that the game has no value except for rare cases. For arbitrary positive integers n and k, we establish an upper bound $0.5n/(2k-1)$ for the lower value of the game and prove its attainability for $k > 1$ and sufficiently large $n = n(k)$. Thus, we show that a clustering of an arbitrary sample of length n can be constructed by the k medians method so that the payoff does not exceed the obtained bound, and the bound is attainable for an arbitrary number of clusters and for sufficiently long samples. These results are applicable in combinatorial optimization in the proof of polynomial solvability of subclasses of intractable extremal problems.

Keywords: clustering, k-medians problem, attainable accuracy guarantee.

REFERENCES

1. Aggarwal C. C., Reddy C. K. *Data clustering: algorithms and applications*. Bosa Roca: Taylor & Francis Inc., 2013, Chapman & Hall/CRC Data Mining and Knowledge Discovery Ser., 652 p. ISBN: 9781466558212 .
2. Ben-David S. Computational feasibility of clustering under clusterability assumptions, CoRR abs/1501.00437, 2015, 27 p. Available at: http://arxiv.org/abs/1501.00437.
3. Duda R. O., Hart P. E., Stork D. G. *Pattern classification*. NY: Wile, 2001, 680 p. ISBN: 978-0-471-05669-0 .
4. Grønlund A., Larsen K. G., Mathiasen A., Nielsen J. S. Fast exact k-means, k-medians and bregman divergence clustering in 1d. CoRR abs/1701.07204, 2017, 12 p. URL: http://arxiv.org/abs/1701.07204 .
5. Guruswami V., Indyk P. Embeddings and non-approximability of geometric problems. *Proc. of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA '03). Philadelphia: Society for Industrial and Applied Mathematics, 2003, pp. 537–538. ISBN: 0-89871-538-5 .
6. Har-Peled S., Mazumdar S. On coresets for k-means and k-median clustering. *Proc. of the Thirty-Sixth Annual ACM Symposium on Theory of Computing* (STOC '04). NY: ACM, 2004, pp. 291–300. doi: 10.1145/1007352.1007400 .
7. Khachay M., Neznakhina K. Generalized pyramidal tours for the generalized traveling salesman problem. *Lecture Notes in Computer Science*, 2017, vol. 10627, pp. 265–277. doi: 10.1007/978-3-319-71150-8-23 .
8. Khachay M., Pankratov V., Khachay D. Attainable best guarantee for the accuracy of k-medians clustering in [0,1]. 8^{th} *International Conf. Optimization and Applications* (OPTIMA2017), Y. G. Evtushenko, et al. eds., Aachen: CEUR Workshop Proceedings, 2017, pp. 322–327.
9. Kumar A., Sabharwal Y., Sen S. Linear-time approximation schemes for clustering problems in any dimensions. *J. ACM. Feb.*, 2010, vol. 57(2), pp. 5:1–5:32. doi: 10.1145/1667053.1667054 .

The paper was received by the Editorial Office on September 22, 2017.

Mikhail Yur'evich Khachai, Dr. Phys.-Math. Sci., Prof., Krasovskii Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences, Yekaterinburg, 620990 Russia; Ural Federal University, Ekaterinburg, 620002 Russia; Omsk State Technical University, Omsk, 644050 Russia, e-mail: mkhachay@imm.uran.ru .

Daniil Mikhailovich Khachai, graduate student, Krasovskii Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, 620990 Russia; Institute of Mathematics and Computer Science, Ural Federal University, Ekaterinburg, 620002 Russia,
e-mail: dmx@imm.uran.ru .

Vasiliy Sergeevich Pankratov, doctoral student, Krasovskii Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, 620990 Russia,
e-mail: pankratov.vs@gmail.com .

Cite this article as:
M. Yu. Khachai, D. M. Khachai, V. S. Pankratov, Attainable best guarantee for the accuracy of k-medians clustering in $[0, 1]$, *Trudy Inst. Mat. Mekh. UrO RAN*, 2017, vol. 23, no. 4, pp. 301–310 .