

An online 1-dimensional clustering problem with variable sized clusters

Gabriella Divéki

In the online clustering problems, the classification of points into sets (called clusters) is done in an online fashion. Points arrive one by one at arbitrary locations, to be assigned to clusters at the time of arrival without any information about the further points. A point can be assigned to an existing cluster, or a new cluster can be opened for it. Existing clusters cannot be merged or split. We study one-dimensional variants. The cost of a cluster is the sum of a fixed setup cost scaled to 1 and the square of the diameter of the cluster. The goal is to minimize the sum of costs of the clusters used by the algorithm.

We study two variants, both maintaining the properties that a point which was assigned to a given cluster must remain assigned to this cluster, and clusters cannot be merged. In the strict variant, the size and the exact location of the cluster must be fixed when it is initialized. In the flexible variant, the algorithm can shift the cluster or expand it, as long as it contains all points assigned to it. We study the online and the semi-online versions of the above two variants.

In [1] the one-dimensional variant of our problem is examined, where there is no restriction on the length of a cluster, and the cost of a cluster is the sum of a fixed setup cost and its diameter. Both the strict and the flexible model have been investigated and an intermediate model, where the diameter is fixed in advance but the exact location can be modified is also studied.

In [2] the two-dimensional problem is considered using the l_∞ norm instead of the l_2 norm. Thus, "balls" are actually squares. The cost of each cluster is a sum of a constant setup cost scaled to 1 and the area of the cluster. The goal is to minimize the total cost of the clusters.

We present the first online algorithms for the solution of the problem. We present algorithms for the strict and the flexible variant. We also give lower bounds on the possible competitive ratio in the case of both of the variants.

References

- [1] J. Csirik, L. Epstein, Cs. Imreh and A. Levin, Online Clustering with Variable Sized Clusters, *In Proceedings of MFCS 2010, LNCS 6281* 282–293, 2010.
- [2] G. Divéki and Cs. Imreh, An online 2-dimensional clustering problem with variable sized clusters, *Submitted for publication, 2011.*