
An adaptive strategy for dynamic data clustering with the K-means algorithm

Giuliano Laccetti¹, Marco Lapegna¹, Valeria Mele¹, Diego Romano²

¹Department of Mathematics and Applications

University of Naples Federico II, Naples, Italy

²Institute for High Performance Computing and Networking (ICAR)

National Research Council (CNR), Naples, Italy

{giuliano.laccetti, marco.lapegna, valeria.mele}@unina.it

diego.romano@cnr.it

K-means algorithm is one of the most widely used methods in data mining and statistical data analysis to partition several objects in K distinct groups, called clusters, on the basis of their similarities. The main problem of this algorithm is that it requires the number of clusters as an input data, but in the real life it is very difficult to fix in advance such value. For such reason, several modified K-means algorithms are proposed where the number of clusters is defined at run time, increasing its number at each iteration until a given cluster quality metric is satisfied. In order to face the high computational cost of this approach we propose an adaptive procedure, where at each iteration two new clusters are created, splitting only the one with the worst value of the quality metric.

Keywords: Modified K-means clustering, adaptive algorithm, unsupervised learning, data mining.