ICM 1998, Berlin, Aug. 18–27

Abstracts of Short Communications and Posters

Section:

15. Numerical Analysis and Scientific Computing

1991 MS Classification: 68, 90
Reinhardt, Gerd, Weierstraß- Institut für Angewandte Analysis und Stochastik, Mohrenstr. 39, 10117 Berlin (http://www.wias-berlin.de)

Schnelle Algorithmen zum Clustern großer Datenmengen

Das Poster zeigt das Leistungsverhalten mehrerer Modifikationen zweier Basisalgorithmen aus dem Gebiet der partitionierenden Clusteranalyse. Die zwei Basisalgorithmen sind (grob klassifiziert) der L_2- und L_1-Algorithmus, deren Unterschied in der Wahl des Distanzmaßes (quadrierte L_2-Norm / L_1-Norm) und der zugehörigen Representanten (Mittelwert / Median) liegt. Durch einfache algorithmische Verbesserungen wird das Leistungsverhalten des L_1-Algorithmus mit dem des L_2-Algorithmus vergleichbar.
Modifikationen der Algorithmen sind bez. der Bestimmung der Representanten möglich. Beim Wechsel eines Objektes in ein anderes Cluster kann nach einer globalen oder einer unmittelbaren Aktualisierung der Representanten unterschieden werden. Beschleunigungen der Algorithmen sind durch Reduzierung der Anzahl der Operationen und/oder durch Anpassung an die Hardware (Speicherhierarchie; Mehrprozessor-Maschinen) möglich.