Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications

Jörg Sander, Martin Ester, Hans-Peter Kriegel, Xiaowei Xu

Institute for Computer Science, University of Munich
Oettingenstr. 67, D-80538 München, Germany
{sander | ester | kriegel | xwxu}@informatik.uni-muenchen.de

Abstract. The clustering algorithm DBSCAN relies on a density-based notion of clusters and is designed to discover clusters of arbitrary shape as well as to distinguish noise. In this paper, we generalize this algorithm in two important directions. The generalized algorithm - called GDBSCAN - can cluster point objects as well as spatially extended objects according to both, their spatial and their non-spatial attributes. In addition, four applications using 2D points (astronomy), 3D points (biology), 5D points (earth science) and 2D polygons (geography) are presented, demonstrating the applicability of GDBSCAN to real world problems.

Keywords: Clustering Algorithms, Spatial Databases, Efficiency, Applications.