Spatial Data Mining: Database Primitives, Algorithms and Efficient DBMS Support

Martin Ester, Alexander Frommelt, Hans-Peter Kriegel, Jörg Sander

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

Abstract: Spatial data mining algorithms heavily depend on the efficient processing of neighborhood relations since the neighbors of many objects have to be investigated in a single run of a typical algorithm. Therefore, providing general concepts for neighborhood relations as well as an efficient implementation of these concepts will allow a tight integration of spatial data mining algorithms with a spatial database management system. This will speed up both, the development and the execution of spatial data mining algorithms. In this paper, we define neighborhood graphs and paths and a small set of database primitives for their manipulation. We show that typical spatial data mining algorithms are well supported by the proposed basic operations. For finding significant spatial patterns, only certain classes of paths "leading away" from a starting object are relevant. We discuss filters allowing only such neighborhood paths which will significantly reduce the search space for spatial data mining algorithms. Furthermore, we introduce neighborhood indices to speed up the processing of our database primitives. We implemented the database primitives on top of a commercial spatial database management system. The effectiveness and efficiency of the proposed approach was evaluated by using an analytical cost model and an extensive experimental study on a geographic database.