

An extended ID3 decision tree algorithm for spatial data

ABSTRACT

Utilizing data mining tasks such as classification on spatial data is more complex than those on non-spatial data. It is because spatial data mining algorithms have to consider not only objects of interest itself but also neighbours of the objects in order to extract useful and interesting patterns. One of classification algorithms namely the ID3 algorithm which originally designed for a non-spatial dataset has been improved by other researchers in the previous work to construct a spatial decision tree from a spatial dataset containing polygon features only. The objective of this paper is to propose a new spatial decision tree algorithm based on the ID3 algorithm for discrete features represented in points, lines and polygons. As in the ID3 algorithm that use information gain in the attribute selection, the proposed algorithm uses the spatial information gain to choose the best splitting layer from a set of explanatory layers. The new formula for spatial information gain is proposed using spatial measures for point, line and polygon features. Empirical result demonstrates that the proposed algorithm can be used to join two spatial objects in constructing spatial decision trees on small spatial dataset. The proposed algorithm has been applied to the real spatial dataset consisting of point and polygon features. The result is a spatial decision tree with 138 leaves and the accuracy is 74.72%.

Keyword: ID3 algorithm; Spatial decision tree; Spatial information gain; Spatial measure; Spatial relation