Analyses of indexing techniques on uncertain data with high dimensionality

ABSTRACT

Deploying a solution for handling critical decision-based problem efficiently requires the processing of high-dimensional data. Over the years, due to modern technological advancement, unprecedented volume of uncertain data is been captured and this has necessitated the need to organize such data for better data access performance. To this effect, the use of indexing technique for supporting, organizing, and storing of uncertain data with high dimensionality has become pertinent. However, the choice of an indexing technique to improve search performance is highly influenced by the properties of the underlying data set, data construction methods employed by the indexing structure, and the query types it supports. This paper is motivated to conduct an extensive performance analysis among existing indexing techniques, namely: R-tree, R*-tree and X-tree, in order to realize the most efficient indexing structure for organizing, storing and ultimately improving search performance over uncertain data with high dimensionality. The results of the analyses with regard to CPU processing time and number of nodes visited clearly show the superiority of X-tree over R-tree and R*-tree, as its superiority holds for different data set sizes, data distributions, number of dimensions and even with varying selectivity ratio.

Keyword: Data partitioning; Indexing techniques; MBR; Uncertain data; High-dimensional data