

Review:An Optimal and Progressive Algorithm for Skyline Queries

Hyountaek Yong

IDS Lab. Postech.

As skyline queries can be effectively identify interesting candidates with low formulation overhead, efficient skyline query algorithms have been widely studied. In particular, second memory based algorithms have received attention as skyline queries are conducted in large databases. I/O access is the bottleneck of second memory base systems so these algorithms uses index such as bitmap, B-tree, R-tree for reducing I/O accesses. Among of them, NN(nearest neighbor) algorithm shows good features such as high speed, progressiveness(returns the results gradually) and fairness, but it has some disadvantages like overlapping partition overhead in higher dimensions and accesses to the same node of R-tree.

This paper proposes BBS(Branch and Bounding Search) algorithm, it overcomes NN's problem by one pass R-tree index traversing method. While NN traverses R-tree index in multiple times to the divided regions after finding the nearest neighbor object, BBS proceeds the traversing after finding nearest neighbor object for finding next nearest neighbor object. Moreover, as BBS prunes dominated nodes in traversing process, it does not access to dominated regions. Therefore, BBS does not access overlapping regions twice and has no space overhead for storing divided regions. Especially, BBS considerably outperforms NN in I/O accesses by single traversing R-tree index, and this paper proves I/O optimality of BBS.

In experiment result, BBS shows good performance compare to NN for both progressive and complete skyline computation. Specially, the performance gap with NN becomes larger by increasing dimensionality.

BBS can be applied variations of skyline query such as ranked skyline query and constrained skyline query, dynamic skyline query, k -dominating query.

However, BBS have two limitations. First, BBS is R-tree index base algorithm while R-tree index is not builded in most of databases. Moreover, index structure can not be supported in combination of queries such as join and skyline query. Second, BBS does not efficiently support k -dominating query. The computing method for exclusive dominating region(EDR) suggested in the paper is not correct in higher dimensions and the correct method is very expensive.¹

BBS can be extended skyline queries on other databases. For example, we can obtain probable skyline results uncertain databases by using probability instead of minimum distance as a criterion which determines a tree traversal order .

¹ The k -dominating problem is solved by *DeltaSky* algorithm.