

IDSa: an efficient algorithm for skyline queries computation on dynamic and incomplete data with changing states

ABSTRACT

Skyline queries have been widely used as an effective query tool in many contemporary database applications. The main concept of skyline queries relies on retrieving the non-dominated tuples in the database which are known skylines. In most database applications, the contents of the databases are dynamic due to the continuous changes made towards the database. Typically, the changes in the contents of the database occur through data manipulation operations (INSERT and/or UPDATE). Performing these operations on the database results in invalidating the most recent skylines before changes are made on the database. Furthermore, the presence of incomplete data in databases becomes frequent phenomena in recent database applications. Data incompleteness causes several challenges on skyline queries such as losing the transitivity property of the skyline technique and the test dominance process between tuples being cyclic. Reapplying skyline technique on the entire updated incomplete database to determine the new skylines is unwise due to the exhaustive pairwise comparisons. Thus, this paper proposes an approach, named Incomplete Dynamic Skyline Algorithm (IDSa) which attempts to determine the skylines on dynamic and incomplete databases. Two optimization techniques have been incorporated in IDSa, namely: pruning and selecting superior local skylines. The pruning process attempts to exploit the derived skylines before the INSERT/UPDATE operation made on the database to identify the new skylines. Moreover, selecting superior local skylines process assists in further eliminating the remaining non-skylines from further processing. These two optimization techniques lead to a large reduction in the number of domination tests due to avoiding re-computing of skylines over the entire updated database to derive the new skylines. Extensive experiments have been accomplished on both real and synthetic datasets, and the results demonstrate that IDSa outperforms the existing solutions in terms of the number of domination tests and the processing time of the skyline operation.

Keyword: Dynamic database; Incomplete database; Pairwise comparison; Skyline queries