The Conceptual Framework of Knowledge of Large Scale and Incomplete Graphs of Skyline Queries Optimization Using Machine Learning

ICETAS 2024 - 9th IEEE International Conference on Engineering Technologies and Applied Sciences • Conference Paper • 2024 • DOI: 10.1109/ICETAS62372.2024.11119942

Noor, Ubair ; Hassan, Raini ; Handayani, Dini Oktarina Dwi

Kulliyah of Information and Communication Technology, International Islamic University Malaysia, Department of Computer Science, Kuala Lumpur, Malaysia

Abstract

Ever since its introduction into the database community, skyline queries have been widely adopted in a range of contemporary database applications. Skyline technique relies on the concept of Pareto-optimal in which a data item from the set of dataset D is identified as skyline if and only if it is not worse than other data items in all dimensions (attributes) and strictly better in at least one dimension. Most of the previous skyline solutions have been designed for conventional databases for complete, incomplete, and uncertain data. However, not much attention has been paid to issues related to skyline query processing over knowledge of large-scale incomplete graph databases. Most recently, graphs have become prevalent data structures to model complex information networks for various real-life contemporary applications such as social networks, knowledge bases, pattern recognition, and the World Wide Web. It is also important to note that, generally graphs are big

structures with very big data and this change often due to updates. These continuous updates makes the graph to be highly dynamic, where nodes/edges are added to or removed from the graph always. However, the issue of data incompleteness when processing skyline queries in large-scale graph databases has not been considered by previous works. The research aims at proposing a new model for processing skyline queries in an incomplete graph database. The research methodology includes reviewing the related literature of skyline queries in incomplete graph databases. Then, propose an method for handling skyline queries within an incomplete graph database followed by designing and implementing a model to evaluate the efficiency and effectiveness of the proposed approaches. The preliminary results using the K means Clustering Algorithm showed that the conceptual framework successfully grouped similar data points, facilitating the identification of skyline points. The implemented algorithm to perform such operation was far more efficient, faster and accurate as compared to conventional methods. This research will ultimately benefit a wide range of applications involving decision-making, decision support, social network, and recommendations aspects by developing a tool that incorporates the proposed approaches. © 2024 IEEE.

Author keywords

Graph Database; Machine Learning; Skyline Query

Indexed keywords

Engineering controlled terms

Behavioral research; Big data; Clustering algorithms; Complex networks; Decision making; Decision support systems; Graph structures; Graphic methods; Knowledge graph; Learning systems; Machine learning; Pattern recognition; Query languages; Query processing

Engineering uncontrolled terms

Conceptual frameworks; Data items; Database applications; Database community; Graph database; Large-scales; Machine-learning; Pareto-optimal; Queries optimization; Skyline query

Engineering main heading

Graph Databases

Funding details

Details about financial support for research, including funding sources and grant numbers as provided in academic publications.

Funding sponsor	Funding number	Acronym
Ministry of Higher Education, Malaysia See opportunities by MOHE		MOHE

Funding text

The Fundamental Research Grant Scheme (FRGS) with the Reference Code support this research FRGS/1/202/ICT01/ UIAM/02/2 from the Ministry of Higher Education (MOHE) Malaysia.