

Learning Database Optimization Techniques: the State of the Art and Prospects

**Shaojie QIAO, Hanlin FAN, Nan HAN, Lan DU, Yuhan
PENG, Rongmin TANG, Xiao QIN**

Frontiers of Computer Science, DOI: [10.1007/s11704-025-41116-7](https://doi.org/10.1007/s11704-025-41116-7)

Problems & Ideas

- Problems of conventional stereo matching approaches:

Traditional database systems struggle with optimizing queries in the era of big data. They fail to handle the complexities of large-scale, multi-table queries and high-dimensional data. The current approaches, including cardinality/cost estimation and join order selection, often lead to suboptimal execution plans due to limited adaptability.

- Ideas:

This paper proposes a shift toward AI-enhanced optimization using machine learning (ML) and deep reinforcement learning (DRL) techniques. The key idea is to improve cardinality/cost estimation, select optimal join orders, optimize Text-to-SQL methods, design end-to-end optimizers by leveraging AI models. This novel approach, especially using deep learning and reinforcement learning, promises more accurate predictions, greater adaptability, and efficiency, overcoming traditional database system limitations.

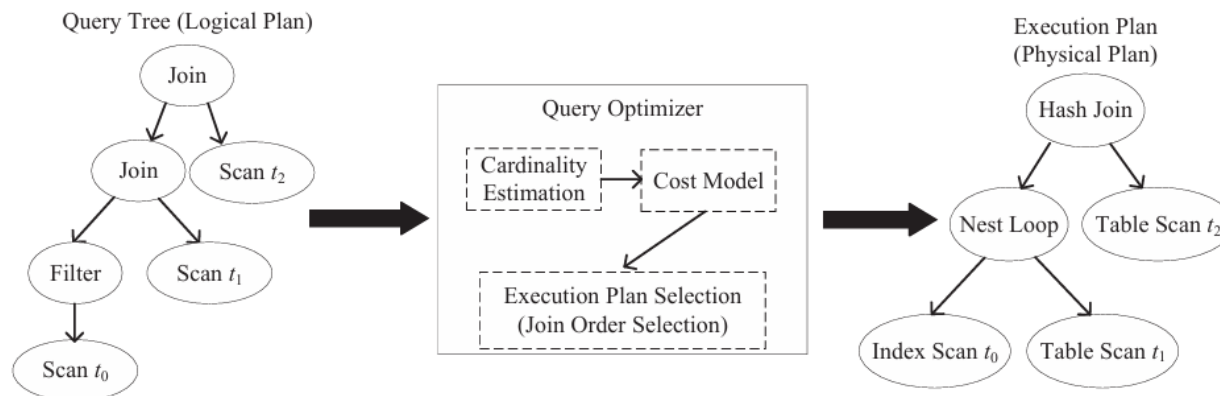


Fig. 2: Overview of a query optimizer

The diagram shows three important components of the cost-based query optimizer: execution plan selection, and cardinality/cost estimation models.

Main Contributions

- Contributions :
 - Machine learning techniques, particularly deep learning and reinforcement learning, offer significant improvements in database optimization.
 - Future directions include the integration of AutoML, online learning, and hybrid models for more adaptive and efficient query optimization.
 - Text-to-SQL models are evolving rapidly, with data augmentation and advanced architectures enhancing their robustness and accuracy.

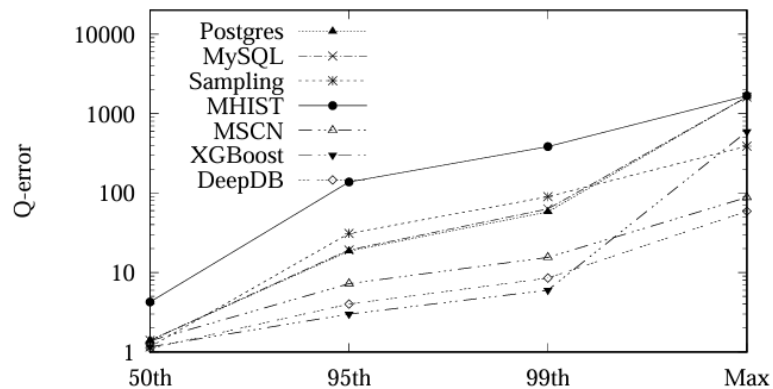


Fig. 9: Q-error comparison of cardinality estimation methods.

The experimental results demonstrate that learning-based methods consistently outperform traditional methods in cardinality estimation and query optimization accuracy.

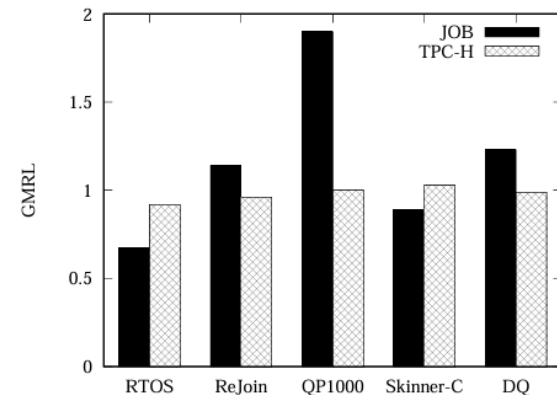


Fig. 11: GMRL comparison of join order selection methods.

The experimental results shows the effectiveness of learning-based models in optimizing query execution, especially for complex queries.