

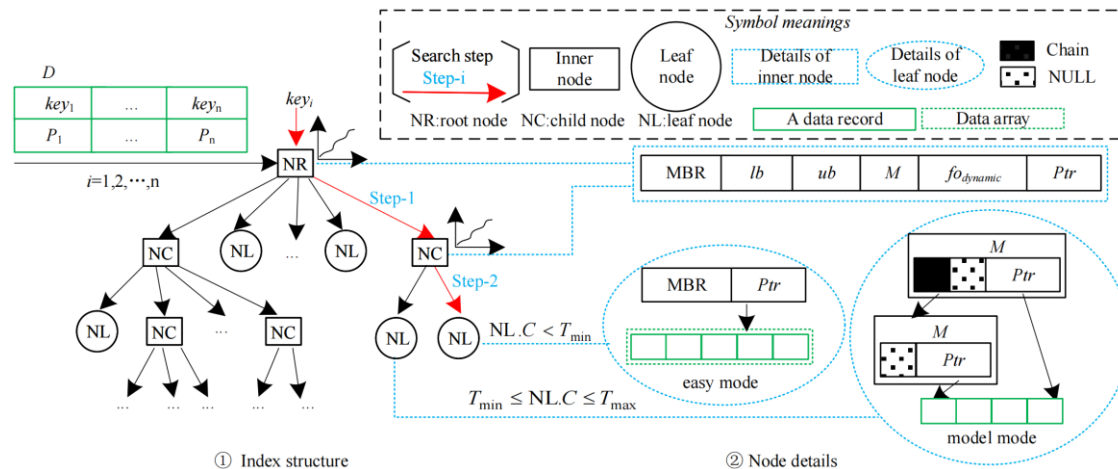
FALSI: a Fast-constructing and Adaptive Learned Spatial Index

Pan CAI, Chaohong MA, Cuiping LI, Hong CHEN

Frontiers of Computer Science, DOI: [10.1007/s11704-025-50868-1](https://doi.org/10.1007/s11704-025-50868-1)

Problems & Ideas

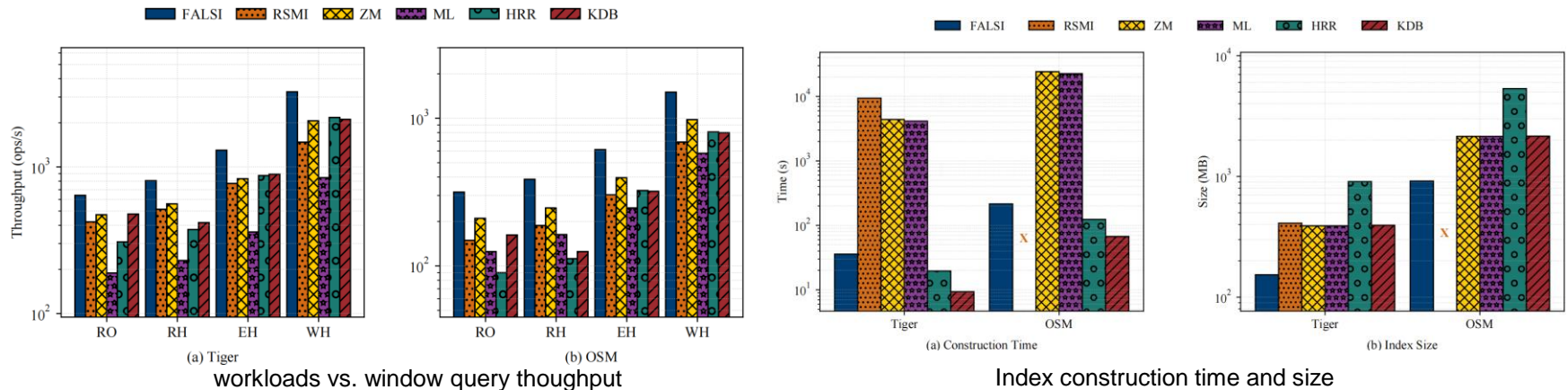
- Problems of learned spatial indexes:
 - prohibitively expensive construction costs and significant update overhead from model retraining.
 - suboptimal query performance due to approximation error.
- Ideas: we propose FALSI, a Fast-constructing Adaptive and Learned Spatial Index, which introduces a novel index structure designed to systematically enhance construction speed, query efficiency, and dynamic update adaptability.



FALSI's core design consists of two key components: First, FALSI implements an autonomous partitioning strategy that utilizes lightweight linear regression models to recursively partition and reorganize spatial data in a top-down manner. Second, we introduce a multi-granularity node optimization strategy that dynamically adjusts node fanout and structural types based on evolving data distribution and predefined threshold.

Main Contributions

- Contributions:
 - An adaptive learned spatial index (FALSI) is proposed to efficiently process mixed read-write workloads under complex data distributions
 - An autonomous partitioning strategy recursively divides data into tree-structured partitions using models until leaf nodes meet a size threshold. The resulting index tree consists of independent subtrees, each rooted by a linear model governing its leaf nodes
 - The multi-granularity node optimization strategy dynamically adjusts node fanout, employing coarse-grained partitioning at upper levels and fine-grained partitioning at lower levels to significantly improve query and insertion performance under complex data distributions



Left: the average throughput of FALSI and baseline methods under workloads with varying read-write ratios, where read operations are window queries. Right: the construction time and index size of the FALSI index and baseline methods