Algorithm For Spatial Join Operations

Presented By :
Jay Vora (M220249CS)

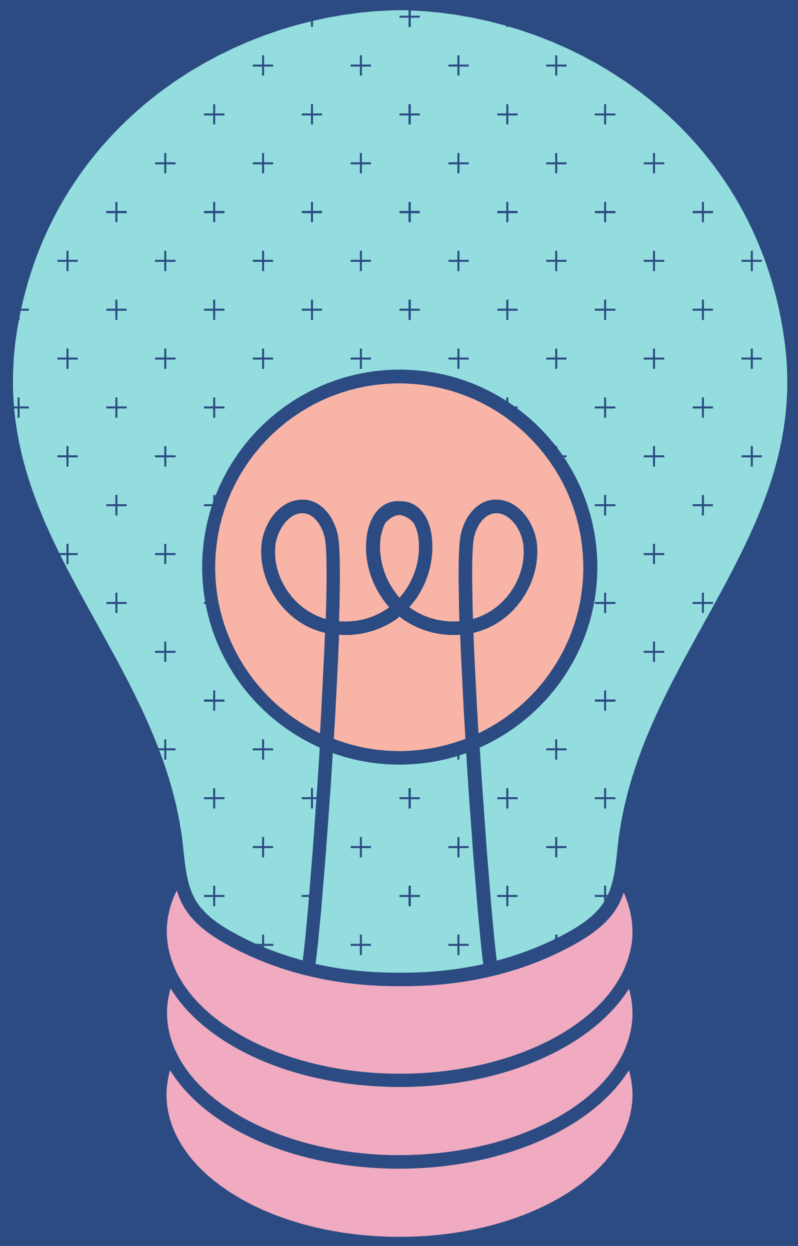
Faculty Incharge :
Dr. S D Madhukumar

Motivation

Importance in GIS: Spatial join is a fundamental operation in Geographic Information Systems (GIS).

Complex algorithms: Spatial join algorithms can be computationally intensive and complex, involving a range of techniques

Real-world examples: Spatial join algorithms can be applied to a wide range of real-world scenarios, providing concrete and relatable examples for learners.



INTRODUCTION

WHAT ARE JOIN QUERIES?

A spatial join involves matching rows from the Join Features to the Target Features based on their relative spatial locations
Example: Find all the forests that are located within any city.

As regular join operation, spatial join is also a time-consuming operation. There are many algorithms which are put forth to reduce the time complexity of these important types of queries.

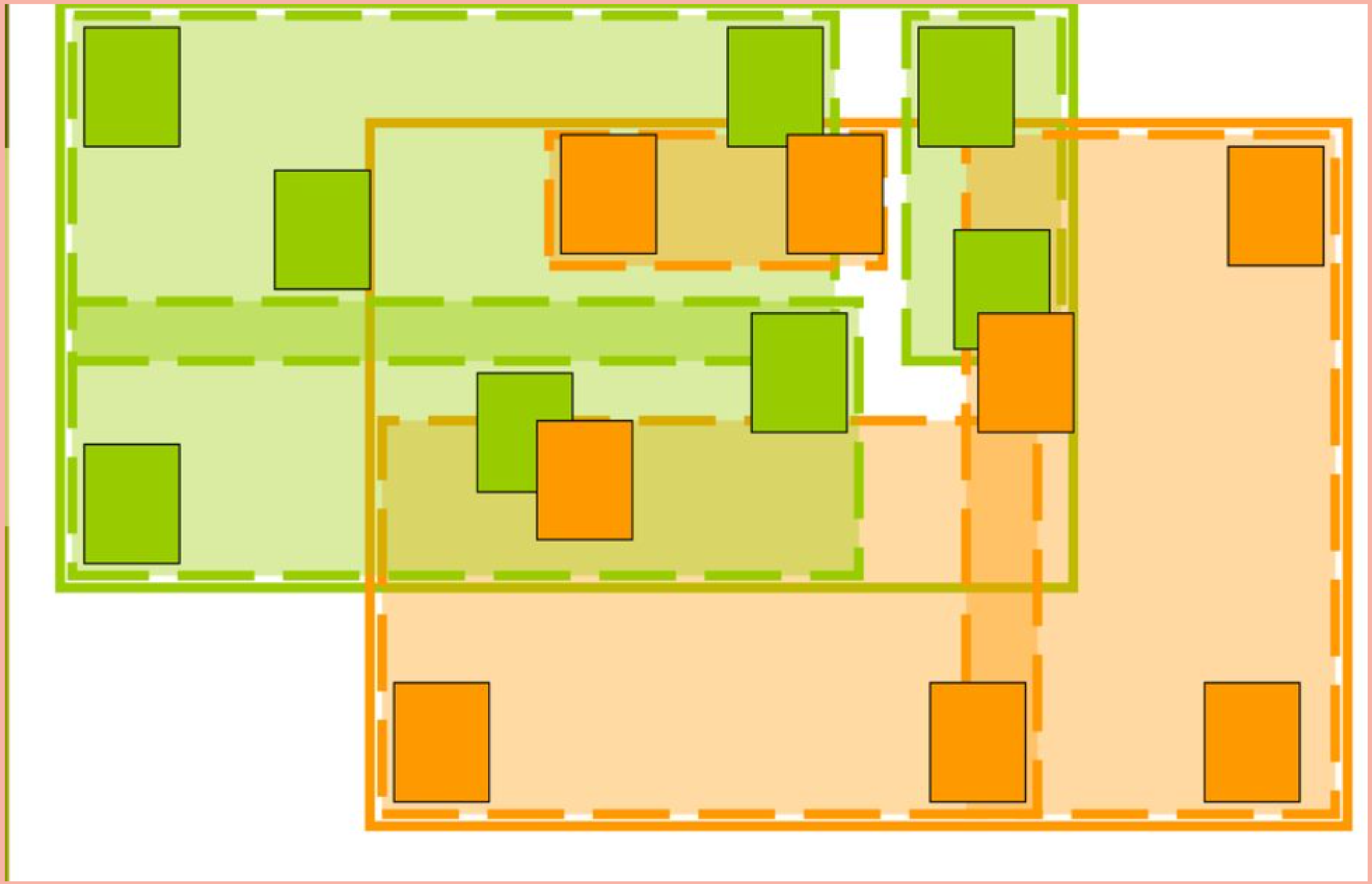
VARIOUS TYPES OF SPATIAL JOIN TECHNIQUES

Internal Memory Methods	1. Nested Loop Join 2. Plane Sweep 3. Z-Order
External Memory Methods Both Sets Indexed	1. Hierarchical Traversal 2. Non-Hierarchical Method 3. Multi-dimensional point methods
External Memory Methods One Set Not Indexed	1. Construct a Second Index 2. Index as Partitioned Data 3. Index as Sorted Data
External Memory Methods Neither Sets Indexed	1. External Plane Sweep 2. Partitioning Algorithms

Breadth First -R Tree Join

- Synchronous traversal of two R-trees in a breadth-first traversal.
- In BFRJ, a global optimization is achieved for each level.
- The main idea is that, based on the global optimization, we can take decisions as which nodes need to be joined to each other.
- Both relations are indexed using PMR quadtree
- Performs a synchronized tree traversal at the leaf level.
- To facilitate the breadth-first search, a queue, called SJQ, is used to save elements

Example of R-Tree Join



kNN Query Algorithm

1 — 2 — 3

STEP 1

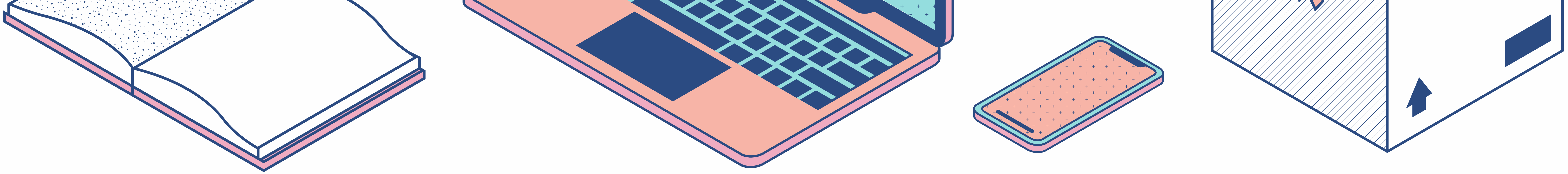
Object Classification

STEP 2

Distance Classification

STEP 3

Concentric Circle Query Region



Object Classification

Object classifiers are objects that indicate their classification for a particular classifier set, and are added to the R-tree by augmenting the node structure with extra data.

Distance Classification

Distance classifiers should generalize very large regions in relation to the amount of the data when used to map a Euclidean distance/increment generator to a specific region. Every sub-regions should utilize the same distance classification as their nearest classified ancestor, and there should be a global distance function.

Concentric Circle Query Region

An initial search circle is generated using a distance classification based on regional characters. When the R-tree is traversed, the algorithm checks whether the current MBB intersects the query circle and that it contains the classifications of the query. If all MBBs have been checked and there are still not k objects found, a function is applied to the current circle to incrementally increase it. If the MBB is contained in the current query circle and not completely contained in the previous query circle, its sub tree is searched and the algorithm can significantly reduce the nodes that it has already searched.

Plane Sweep Algorithm

- The two-dimensional plane-sweep of a set of iso-oriented rectangles that have two passes: the first sorts the rectangles in ascending order, and the second sweeps a vertical scan line from left to right.
- The key to the algorithm is its ability to keep track of the active rectangles and perform the intersection test.
- The plane-sweep algorithm uses three operations to track active rectangles: INSERT, REMOVE INACTIVE, and SEARCH.
- A spatial join determines all pairs of intersecting rectangles in A and B, while a plane-sweep algorithm applies a sweep structure for both A and B. The sweep structures' implementation data structure may significantly affect how well they perform as measured by their performance studies.

Partition Based Spatial Merge (PBSM)

1

STEP

Filter Step

2

STEP

Refinement Step

Partition Based Spatial Merge

Filter Step

The PBSM algorithm uses an approximation of the spatial feature to get a ”rough estimate” of the characteristics of the spatial attribute, partitioning large inputs into smaller chunks, and using a computational geometry plane-sweeping technique to join the chunks.

Partition Based Spatial Merge

Refinement Step

As the filter step ”joins” the inputs based on the MBR of the joining attributes, it is possible for two non-overlapping spatial features to have overlapping MBRs. As a result, the filter step typically produces a superset of the joint outcome. The R and S tuples that are represented by the OID pair created in the previous step are retrieved from disc in the refinement stage, and their join attributes are checked to see if the join predicate is indeed satisfied.

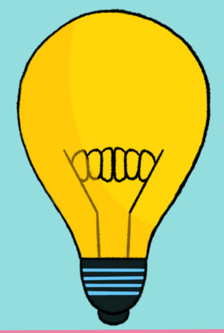
Challenges

- **Complexity:** Spatial join algorithms can be complex, especially when dealing with large datasets.
- **Choice of algorithm:** There are multiple algorithms available for performing spatial join operations, each with its own strengths and weaknesses.
- **Data preparation:** Spatial join algorithms require the input data to be properly prepared and formatted.
- **Visualization:** Spatial join operations can produce complex and multi-dimensional output data, which can be difficult to visualize and interpret.

Conclusion

In conclusion, spatial join algorithms are an essential tool in geographic information systems (GIS) and data analysis. They allow us to combine data from multiple sources based on their spatial relationships and provide valuable insights into complex spatial patterns.

There are various spatial join algorithms available, each with its own strengths and weaknesses. The choice of algorithm depends on the nature of the data, the desired output, and the computational resources available.



References

[1] K. Bhima, T. A. Sri, K. D. Ramaiah, and A. Jagan, “Exerting spatial join and knn queries on spatial database,” in 2012 International Conference on Recent Advances in Computing and Software Systems. IEEE, 2012, pp. 260–266.

[2] E. H. Jacox and H. Samet, “Spatial join techniques,” ACM Transactions on Database Systems (TODS), vol. 32, no. 1, pp. 7–es, 2007.

[3] T. Vu, A. Belussi, S. Migliorini, and A. Eldawy, “A learned query optimizer for spatial join,” in Proceedings of the 29th International Conference on Advances in Geographic Information Systems, 2021, pp. 458–467.

Thank you