Learned Spatial Indexes

Keywords: Spatial index, Machine learning

Problem:
Spatial indexes such as R-trees and quadtrees are used to efficiently search for spatial objects like points or polygons in spatial databases. Such indexes are crucial in optimizing the performance of spatial queries. However, existing spatial indexes are oblivious to the distribution and patterns of objects in space and treat all objects and regions in space alike.

Recent work in database systems aims to use learned indexes [1,2] for databases. Learned indexes aim to improve data access times over traditional database indexes by learning data distributions and data access patterns for the data. A spatial index, however, indexes different types of spatial objects like points, lines and polygons over more than one dimension. Existing strategies for learned indexes hence may not be suitable for a spatial index structure.

Project: The goal of the project is to develop learned spatial index based on machine learning models. The spatial index would learn the data distributions of the underlying spatial objects and aim to predict the position of the object being searched. We will evaluate the index based on how much time it takes to learn the index, the amount of storage space it takes to store the index and the time taken to access data using the index as compared to existing spatial indexes. We also plan on considering inserts, deletes and updates on the underlying data.

Plan:

1. A literature survey of existing learned indexes and spatial indexes
2. Implement machine learning models to train spatial index structure
3. Evaluate the learned spatial index against existing spatial index structures

Supervisor: Prof. Anastasia Ailamaki, anastasia.ailamaki@epfl.ch
Responsible collaborator(s): Bikash Chandra, bikash.chandra@epfl.ch
Duration: 6-12 months

[1] T. Kraska, A. Beutel, E. H. Chi, J. Dean, and N. Polyzotis. The Case for Learned Index Structures. In SIGMOD, 2018.
[2] A Llaveshi, U Sirin, A Ailamaki and R West. Accelerating B+tree Search by Using Simple MachineLearning Techniques. In AIDB 2019