SPEAKER: Tobias Janssenn

TITLE: Parallel Algorithm for Nearest Neighbor Search

ABSTRACT

The Nearest Neighbor Problem is to design a data structure for a large set S of points in (say) n-dimensional space such that the following problem can be solved quickly: given a point p in n-dim space, find the point in S that is closest to p. The main concerns are that the data structure does not take too much space, and that the nearest-neighbor query be fast.

We vary this problem in two ways: (1) rather than search for the nearest neighbor we will settle for an Approximate Nearest Neighbor ANN, (2) the algorithm that finds the ANN is a parallel algorithm.

One application of ANN is that n-dim datasets are becoming omnipresent due to the widespread adoption of deep learning methods which represent complex objects as high-dime vectors called vector embeddings, which can span thousands of dimensions.