Data Structures and Algorithms for Nearest Neighbor Search in General Metric Spaces

Peter N. Yianilos

Abstract: We consider the computational problem of finding nearest neighbors in general metric spaces. Of particular interest are spaces that may not be conveniently embedded or approximated in Euclidian space, or where the dimensionality of a Euclidian representation is very high.

Also relevant are high-dimensional Euclidian settings in which the distribution of data is in some sense of lower dimension and embedded in the space.

The *vp-tree* (vantage point tree) is introduced in several forms, together with associated algorithms, as an improved method for these difficult search problems. Tree construction executes in O(n log n) time, and search is under certain circumstances and in the limit, O(log n) expected time.

The theoretical basis for this approach is developed and the results of several experiments are reported. In Euclidian cases, kd-tree performance is compared.

Keywords: Metric Space, Nearest Neighbor, Computational Geometry, Associative Memory, Randomized Methods, Pattern Recognition, Clustering.

Links:

- [View the complete paper (converted to HTML)](#)

- [The complete paper (PDF)](#)

- [The complete paper (PostScript)](#)

- [BibTex entry](#)

- My homepage ([Peter N. Yianilos](#))

- See also:
    - [Locally Lifting the Curse of Dimensionality for Nearest Neighbor Search](#)
    - [Excluded Middle Vantage Point Forests for Nearest Neighbor Search](#)