

MSc thesis defense presentation

Dimitrios Nikolopoulos defends his MSc thesis

Date:	Wednesday, 19 Mar 2014
Thesis title:	Randomly-oriented RKD-trees
Committee:	<ul style="list-style-type: none">• Ioannis Emiris• D. Gounopoulos• Manolis Koubarakis

Thesis abstract

Consider a set S of points in a real D -dimensional space \mathbb{R}^D , where distances are defined using function $\Delta : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ (the Euclidean metric). Nearest neighbor search is an optimization problem for finding the closest points in S to a given query point $q \in \mathbb{R}^D$. Given a positive real $\epsilon > 0$ then a point $p \in S$ is a $(1+\epsilon)$ -approximate nearest neighbor of the query point $q \in \mathbb{R}^D$ if $\text{dist}(q, p) \leq (1 + \epsilon)\text{dist}(q, p_{nn})$ where $p_{nn} \in S$ is the true nearest neighbor to q . If the data that is expressed in high- dimensional space \mathbb{R}^D lies closer to an embedded manifold M of dimension d , where $d \ll D$, then, we show the data may be preprocessed into the Randomly-oriented RKD-trees structure and we provide a near optimal bound on the number of levels required to reduce the size of its cells by a factor $s \geq 2$. We show the data may be preprocessed into the structure in $O(D \cdot N \cdot \log N)$ time and $O(D \cdot N)$ space, so that given a query point $q \in \mathbb{R}^D$ and $\epsilon > 0$, a $(1+\epsilon)$ -approximate nearest neighbor of q may be found in high-dimensional data with an underlying low-intrinsic dimensional subspace.

Download date: 2024-10-15, 18:47.