Abstract

We present a framework for similarity search based on Locality-Sensitive Filtering (LSF), generalizing the Indyk-Motwani (STOC 1998) Locality-Sensitive Hashing (LSH) framework to support space-time tradeoffs. Given a family of filters, defined as a distribution over pairs of subsets of space that satisfies certain locality-sensitivity properties, we can construct a dynamic data structure that solves the approximate near neighbor problem in d-dimensional space with query time $dn^{\rho_q + o(1)}$, update time $dn^{\rho_u + o(1)}$, and space usage $dn + n^{1+\rho_u+o(1)}$ where n denotes the number of points in the data structure. The space-time tradeoff is tied to the tradeoff between query time and update time (insertions/deletions), controlled by the exponents ρ_q, ρ_u that are determined by the filter family.

Locality-sensitive filtering was introduced by Becker et al. (SODA 2016) together with a framework yielding a single, balanced, tradeoff between query time and space, further relying on the assumption of an efficient oracle for the filter evaluation algorithm. We extend the LSF framework to support space-time tradeoffs and through a combination of existing techniques we remove the oracle assumption.

Laarhoven (arXiv 2015), building on Becker et al., introduced a family of filters with space-time tradeoffs for the high-dimensional unit sphere under inner product similarity and analyzed it for the important special case of random data. We show that a small modification to the family of filters gives a simpler analysis that we use, together with our framework, to provide guarantees for worst-case data. Through an application of Bochner's Theorem from harmonic analysis by Rahimi & Recht (NIPS 2007), we are able to extend our solution on the unit sphere to $\mathbb{R}^d$ under the class of similarity measures corresponding to real-valued characteristic functions. For the characteristic functions of s-stable distributions we obtain a solution to the (r, cr)-near neighbor problem in ℓ_s^d-spaces with query and update exponents $\rho_q = \frac{c^s(1+\lambda)^2}{(c^s+\lambda)^2}$ and $\rho_u = \frac{c^s(1-\lambda)^2}{(c^s+\lambda)^2}$ where $\lambda \in [-1, 1]$ is a tradeoff parameter. This result improves upon the space-time tradeoff of Kapralov (PODS 2015) and is shown to be optimal in the case of a balanced tradeoff, matching the LSH lower bound by O'Donnell et al. (ITCS 2011) and a similar LSF lower bound proposed in this paper. Finally, we show a lower bound for the space-time tradeoff on the unit sphere that matches Laarhoven's and our own upper bound in the case of random data.