Hardness and Approximation of High-Dimensional Search Problems

Rasmus Pagh*

IT University of Copenhagen, Denmark
pagh@itu.dk

Abstract

The need to perform search in a collection of high-dimensional vectors arises in many areas of computer science including databases, image analysis, information retrieval, and machine learning. In contrast to lower-dimensional settings, we do not know of worst-case efficient data structures for such search problems. In fact, if we make no assumptions on the input there is no known way of doing significantly better than brute force. In this talk I survey recent developments in the theoretical study of high dimensional search problems, including:

- Conditional hardness results linking search problems to well-known computationally hard problems such as k-SAT.
- Upper bounds for *approximate* high-dimensional search using locality-sensitive maps and filters, and work towards derandomizing these algorithms.
- Surprising upper bounds in *batched* settings where there are many simultaneous searches.

The talk ends by sketching directions for future research, connecting to other areas of theoretical computer science but also attempting to obtain better theoretical models that explain the performance of search algorithms that are used in practice.

1998 ACM Subject Classification E.1 Data Structures, F.2.2 Nonnumerical Algorithms and Problems – Geometrical problems and computations

Keywords and phrases Similarity search, hardness, approximation

Digital Object Identifier 10.4230/LIPIcs.MFCS.2017.83

Category Invited Talk

* Supported by the European Research Council under the European Union's 7th Framework Programme (FP7/2007-2013) / ERC grant agreement no. 614331.