Beyond P vs. NP: Quadratic-Time Hardness for Big Data Problems

Piotr Indyk
CSAIL, MIT
indyk@mit.edu

ABSTRACT

The theory of NP-hardness has been very successful in identifying problems that are unlikely to be solvable in polynomial time. However, many other important problems do have polynomial time algorithms, but large exponents in their time bounds can make them run for days, weeks or more. For example, quadratic time algorithms, although practical on moderately sized inputs, can become inefficient on big data problems that involve gigabytes or more of data. Although for many problems no sub-quadratic time algorithms are known, any evidence of quadratic-time hardness has remained elusive.

In this talk I will give an overview of recent research that aims to remedy this situation. In particular, I will describe hardness results for problems in string processing (e.g., edit distance computation or regular expression matching) and machine learning (e.g., Support Vector Machines or gradient computation in neural networks). All of them have polynomial time algorithms, but despite extensive amount of research, no near-linear time algorithms have been found for many variants of these problems. I will show that, under a natural complexity-theoretic conjecture, such algorithms do not exist. I will also describe how this framework has led to the development of new algorithms.

CCS Concepts

Theory of computation

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SPAA'17, July 24-26, 2017, Washington, DC, USA.

ACM ISBN 978-1-4503-4593-4/17/07.
DOI: http://dx.doi.org/10.1145/3087556.3087603

Author Keywords

Edit Distance; Regular Expression Matching; Support Vector Machines; Neural Networks; Fine-grained Complexity

BIOGRAPHY

Piotr Indyk is a Professor of Electrical Engineering and Computer Science at MIT. He joined MIT in 2000, after earning PhD from Stanford University. Earlier, he received Magister degree from Uniwersytet Warszawski in 1995. Piotr's research interests lie in the design and analysis of efficient algorithms. Specific interests include high-dimensional computational geometry, streaming and sub-linear algorithms, sparse recovery and machine learning. He is an ACM Fellow. His work on Sparse Fourier Transform has been named to Technology Review "TR10" in 2012, while his work on locality-sensitive hashing has received the 2012 Kanellakis Theory and Practice Award..

REFERENCES

1. A. Backurs, P. Indyk, "Edit distance cannot be computed in strongly sub-quadratic time (unless SETH is false)", STOC 2015.
2. A. Backurs, P. Indyk, "Which Regular Expression Patterns are Hard to Match?", FOCS 2016.
3. A. Backurs, P. Indyk, L. Schmidt, "On the Fine-Grained Complexity of Empirical Risk Minimization: Kernel Methods and Neural Networks", arXiv:1704.02958, 2017.