

Accepted Manuscript

On the role of statistics in the era of big data: A computer science perspective

Stefano Ceri

PII: S0167-7152(18)30064-6

DOI: <https://doi.org/10.1016/j.spl.2018.02.019>

Reference: STAPRO 8141

To appear in: *Statistics and Probability Letters*



Please cite this article as: Ceri S., On the role of statistics in the era of big data: A computer science perspective. *Statistics and Probability Letters* (2018), <https://doi.org/10.1016/j.spl.2018.02.019>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

On the role of statistics in the era of big data: a computer science perspective

Stefano Ceri

Data Science Group, DEIB, Politecnico di Milano, Milano, Italy

Abstract

Statistics and computer science are facing remarkably similar discussions on the role of big data. In this article, I advocate that the computer science community has taken advantage of big data since about five decades, thereby building the main commercial companies of today's computer industry, and specifically I describe the new emphasis on data as the emergence of the so-called Fourth Paradigm. Then, I draw a parallel between the debates on big data occurring within the statistics and computer science community; and finally I advocate for a joint, new and pervasive approach to data science, in which both communities can capitalize on each other's skills.

Keywords: big data, paradigm shift in statistics and computer science, fourth paradigm, two cultures, teaching data science

1. Big Data in the last five decades

"Big data" is a successful buzzword, used (and abused) to denote a paradigm shift of the scientific approach. As I am a computer scientist, and more specifically a member of database community, I do not consider big data as a revolution. Rather, big data have been the main ingredient of more than five decades of data management research; what happened in recent years is just the natural evolution of what happened in the past.

Download English Version:

<https://daneshyari.com/en/article/7548542>

Download Persian Version:

<https://daneshyari.com/article/7548542>

[Daneshyari.com](https://daneshyari.com)