

2017 Onassis Lectures in Computer Science on Big Data and Applications

The Foundation for Research and Technology-Hellas (FORTH)

Heraklion, Crete, Greece

July 10-15, 2017

Big Data that emerged in the early 21st Century marked a significant shift not just in computing and science but in every human endeavor – in how the world operates – with a new, emerging, more significant role of data and data analytics in all human endeavors. These lectures concern Data Science – the emerging principles and techniques that underlie this new paradigm of discovery.

Conclusions

- “Big Data, Data Analytics, Data Science, AI, ...” is the most significant trend in early 21st Century
- Extremely high expectations to transform almost every human endeavor
- Enormous benefits and risks
- A new accelerated paradigm of discovery
- Data Science is in its infancy and has yet to emerge
- Fundamental challenges abound

Lecturer

Dr. Michael L. Brodie is a research scientist in the Computer Science and Artificial Intelligence Lab at MIT. As Chief Scientist of Verizon, the 2nd largest Telco in the world, for 25 years, he has a keen interest in advanced technology and its applications in the real world. As chairman of the Scientific Advisory Board of two of the world’s 50+ Data Science Research Institutes [[Insight Center for Data Analytics](#), Ireland, (2015-), and [Swinburne Data Science Research Institute](#) (2017-)] he is responsible for understanding the opportunities, state of the art, and research challenges for the emerging discipline of Data Science. His lectures present the Big Picture of Big Data and of Data Science and the consequent revolutions in science and industry.

Data: The World’s Most Valuable Resource

Data is being conceived as having potential for transforming all human endeavors for which adequate data is available. Seven of the world’s largest ten enterprises are data-driven companies, mere startups two decades ago. To compete, corporations are transforming themselves to be data-driven. Based on Big Data and Data Science, science and engineering are entering the 5th *paradigm of discovery*. Yet, Data Science is in its infancy without adequate principles to distinguish correlation from causation. This lecture explores the emergence of Big Data and Data Science.

1 Introduction: Big Data and Data Science

- 1.1 Data: The World’s Most Valuable Resource
- 1.2 Opportunities: Accelerating Discovery
- 1.3 Cautions: What Can We Discover? Is it True?
- 1.4 Threats: Power, Disruption, Revolution

2 Data: The World’s Most Valuable Resource

- 2.1 World’s Most Valuable Resource
- 2.2 Data as an Asset Class: Why Data Marketplaces Failed
- 2.3 Big Data Revolution: Current Indicators
- 2.4 Data Analytics/Data Science: Most Significant Trend in early 21st Century

3 Big Data: History, Definition, and Revolutions

- 3.1 Big Data Before “Big Data”
- 3.2 Big Data History
- 3.3 Big Data Definition
- 3.4 Lesson: Think of Data as (Crime/Empirical) Evidence of a Phenomenon
- 3.5 Related Revolutions: Profound Change

3.6 Lesson: Not “Big Data” But Data Science = Big Data + Algorithms (AI) + Models + ...

Big Data and Data Science: State of the Art and Research Directions

While data analytics has been used since before Pharaonic Egypt, it is now becoming a powerful force in discovery and prediction, notwithstanding domain expertise, e.g., in economics, that economic trends are inherently unpredictable. On the other hand, data science has led to accelerating discovery in many domains, e.g., cancer cures, exoplanets, paleontology, FinTech, and retail optimization. Equally powerful threats abound, e.g., influencing the 2016 US election. This lecture discusses the state of the art of Big Data and Data Science in industry and explores the research that lies ahead in the decade as data science matures.

4 Data Science State of the Art: Use Cases

- 4.1 Big Data and Data Science: State of the Art and Research Directions
- 4.2 Data Analytics Use Cases (in 40 domains)
- 4.3 Big Data / Data Science Value in Science & Industry
- 4.4 Big Data / Data Science Trends

5 Data Science: The 5th Paradigm of Discovery

- 5.1 Data Science History
- 5.2 4th Paradigm of Scientific & Engineering Discovery (eScience, Jim Gray)
- 5.3 The Big Picture: The Emergence of Modern Data Analytics (2005-2025)
- 5.4 5th Paradigm of Discovery (Brodie)
- 5.5 Illustration: Oncology/Cancer Research (Cancer Commons)
- 5.6 Conclusions: Confusion: Don't Believe Everything You Read

6 Core Data Science

- 6.1 It's All About Modelling
- 6.2 Inherent (Critical) Data Science Characteristics
- 6.3 Data Analytics Workflows / Pipelines
- 6.4 Big Data Technology / Infrastructure

7 Defining Data Science: An Emerging Discipline

- 7.1 Definition: What is Data Science (in theory)?
- 7.2 What is Data Science Research?
- 7.3 What is a Data Scientist?

8 Data Science Research Challenges

- 8.1 Data Analytics: Next 10 Years
- 8.2 Fundamental Data Science Challenges