

January 2016 | Volume 104 | Number 1

Proceedings OF THE IEEE

SPECIAL ISSUE

Big Data: Theoretical Aspects

Point of View: The Power
of Systems Biology

Scanning Our Past:
Georg Simon Ohm

SPECIAL ISSUE

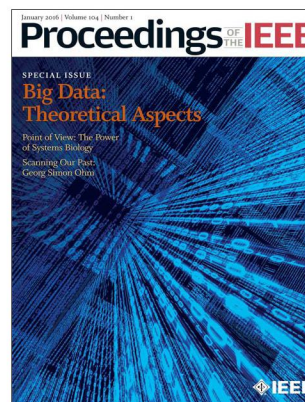
BIG DATA: THEORETICAL ASPECTS

Edited by S. Haykin, S. Wright, and Y. Bengio

- 11 **A Review of Relational Machine Learning for Knowledge Graphs**
By M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich
|INVITED PAPER| This paper reviews how statistical models can be “trained” on large knowledge graphs and then used to predict new facts about the world.
- 34 **Learning to Hash for Indexing Big Data—A Survey**
By J. Wang, W. Liu, S. Kumar, and S.-F. Chang
|INVITED PAPER| This paper provides readers with a systematic understanding of insights, pros, and cons of the emerging indexing and search methods for Big Data.
- 58 **Implementing Randomized Matrix Algorithms in Parallel and Distributed Environments**
By J. Yang, X. Meng, and M. W. Mahoney
|INVITED PAPER| In this paper, the authors review recent work on developing and implementing randomized matrix algorithms in large-scale parallel and distributed environments.
- 93 **Foundational Principles for Large-Scale Inference: Illustrations Through Correlation Mining**
By A. O. Hero, III and B. Rajaratnam
|INVITED PAPER| This paper presents a framework for answering the fundamental question: “When can reliable inference be drawn in the “Big Data” context?”
- 111 **Resource Allocation for Statistical Estimation**
By Q. Berthet and V. Chandrasekaran
|CONTRIBUTED PAPER| Adopting a general view of the notion of a resource and its effect on the quality of a data source, the authors describe in this paper a framework for the allocation of a given set of resources to a collection of sources in order to optimize a specified metric of statistical efficiency.
- 126 **Magging: Maximin Aggregation for Inhomogeneous Large-Scale Data**
By P. Bühlmann and N. Meinshausen
|INVITED PAPER| In this paper, the authors show how maximum aggregation can address certain challenges in large-scale data analysis of inhomogeneous data.
- 136 **Learning Reductions That Really Work**
By A. Beygelzimer, H. Daumé, III, J. Langford, and P. Mineiro
|INVITED PAPER| This paper summarizes the mathematical and computational techniques that have enabled learning reductions to effectively address a wide class of tasks.
- 148 **Taking the Human Out of the Loop: A Review of Bayesian Optimization**
By B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas
|INVITED PAPER| The paper introduces the reader to Bayesian optimization, highlighting its methodical aspects and showcasing its applications.
- 176 **Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets**
By M. K. K. Leung, A. DeLong, B. Alipanahi, and B. J. Frey
|INVITED PAPER| This paper provides an introduction to machine learning tasks that are relevant to important problems in genomic medicine.

DEPARTMENTS

- 2 **POINT OF VIEW**
Solving Puzzles With Missing Pieces: The Power of Systems Biology
By J. T. Yurkovich and B. O. Palsson
- 8 **SCANNING THE ISSUE**
Big Data: Theoretical Aspects
By S. Haykin, S. Wright, and Y. Bengio
- 198 **SCANNING OUR PAST**
Georg Simon Ohm and the First Comprehensive Theory of Electrical Conductivity in Metals
By S. Reif-Acherman
- 210 **FUTURE SPECIAL ISSUES/SPECIAL SECTIONS**



On the Cover: Our cover image this month is an abstract representation of the term “Big Data,” which refers to data sets that are very large and complex.