Preface

Big data is no longer just a buzz word. It is a serious analytics area, requiring a rethinking of platforms, computing paradigms, and architectures. Big data analytics is now quintessential for a wide variety of traditional as well as modern applications, which has orchestrated intense research efforts in recent years.

This volume contains the papers presented at the 4th International Conference on Big Data Analytics (BDA 2015) held during December 15–18, 2015, in Hyderabad, India. The aim of the conference was to highlight recent advancements in the area while stimulating fresh research. There were 61 submissions in the research track. Each submission was reviewed by on average 2.5 Program Committee members. The committee decided to accept nine papers, which are included in this volume. The volume also includes nine invited papers. The volume is divided into four sections: *Security and Privacy, Commerce, Models and Algorithms, and Medicine.*

The section "Big Data: Security and Privacy" includes two papers. Barker describes the current state of the art and makes a call to open a dialog between the analytics and privacy communities. Agarwal et al. discuss online radicalization and civil unrest as two important applications of open source social media intelligence. The authors also discuss open research problems, important papers, publication venues, research results, and future directions.

The section "Big Data in Commerce" includes four papers. Singh discusses information exploration in e-commerce databases and identifies limitations in the query and result panel that deter exploratory search. Add-on extensions are proposed to address these limitations. Reddy proposes a framework to harvest the page views of the Web by forming clusters of similar websites. Anand et al. propose the utility-based fuzzy miner (UBFM) algorithm to efficiently mine a process model driven by a utility threshold. The utility can be measured in terms of profit, value, quantity, or other expressions of user's preference. The work incorporates statistical and semantic aspects while driving a process model. Bansal et al. integrate the notion of discount in state-of-the-art utility-mining algorithms and propose an algorithm for efficiently mining high-utility itemsets.

The section "Models and Algorithms" includes papers on a wide range of algorithms. Bhatnagar emphasizes that a completely new rethink of the solution from the perspective of the powers of the Map-Reduce paradigm can provide substantial gains. He presents an example of the design of a model-learning solution from high-volume monitoring data from a manufacturing environment. Toyoda addresses the requirement to increase the resilience and safety of transportation systems in view of the 2020 Olympic games in Tokyo, amidst predictions of big earthquakes. Based on large-scale smart card data from the Tokyo Metro subway system and vehicle recorder data, the author presents methods to estimate passenger flows, changes in the flows after accidents, and visualizes traffic with social media streams. Kiran and Kitsuregawa review issues involved in finding periodic patterns in the context of time series and

transactional databases. They describe the basic model of finding periodic patterns, and its limitations, and suggest approaches to address them. Bhardwaj and Dash propose a novel density-based clustering algorithm for distributed environments using the MapReduce programming paradigm. Goel and Chaudhary present an incremental approach for mining formal concepts from a context that is assumed to be present in the form of object-attribute pairs. The approach utilizes the Apache Spark framework to discover and eliminate the redundant concepts in every iteration. Sachdev et al. present an implementation of the alpha-miner algorithm in MySQL and Cassandra using SQL and CQL with the aim of conducting a performance benchmarking and comparison of the alpha-miner algorithm on row-oriented and NoSQL column-oriented databases. Nagpal and Gaur propose a novel algorithm to filter out both irrelevant and redundant features from the data set using a greedy technique with gain ratio.

The last section, "Big Data in Medicine," addresses the applications of big data in the area of medicine. Talukder reviews the science and technology of big data translational genomics and precision medicine with respect to formal database systems. The paper presents two big-data platforms iOMICS (deployed at Google cloud) for translational genomics and DiscoverX (deployed at Amazon Web Services) for precision medicine. Adhil et al. present a clinical expert system (CES) that uses the clinical and genomic marker of the patient combined with a knowledge-base created from distributed, dissimilar, diverse big data. The system predicts the prognosis using a cancer registry compiled between 1997 and 2012 in the USA. Agarwal et al. present an integrative analytics approach for cancer genomics that takes the multiscale biological interactions as key considerations for model development. Sharma et al. propose a class-aware exemplar discovery algorithm, which assigns preference value to data points based on their ability to differentiate samples of one class from others. Goel et al. suggest a system for predicting the risk of cardiovascular disease based on electrocardiogram tests. The authors also discuss a recommendation system for suggesting nearby relevant hospitals based on the prediction.

In closing, we wish to thank the supporting institutes: NIT-Warangal, the University of Delhi, University of Aizu, and the Indian Institute of Technology Delhi. Thanks are also due to our sponsors: KDNuggets, NIT Warangal, Tata Consultancy Services, and Springer. We take this opportunity to thank Niladri Chatterjee (Student Symposium Chair), and Ramesh Agrawal (Tutorial Chair) for helping us in organizing the respective tracks. We gratefully acknowledge the Program Committee members and external reviewers for their time and diligent reviews. Thanks are due to Steering Committee members for their guidance and advisory role. The Organizing Committee and student volunteers of BDA 2015 deserve special mention for their support.

Last but not the least, we appreciate and acknowledge the EasyChair conference management system, which transformed the tedious task of managing submissions into an enjoyable one.

December 2015   Naveen Kumar
                                                                   Vasudha Bhatnagar