Special Section on the International Conference on Data Engineering 2016

Meichun Hsu, Alfons Kemper, and Timos Sellis, Fellow, IEEE

This special section covers extended versions of top-ranked papers of the 32nd International Conference on Data Engineering that was held in Helsinki, Finland, May 16-May 20, 2016. The ICDE 2016 conference received 395 submissions within the research track out of which 99 research papers were accepted. Within the extremely well-structured reviewing process, every paper was at least reviewed by three members of the program committee (PC). The PC consisted of 155 members; this year, special emphasis was laid on a balanced structure of the committee by appointing experienced senior researchers as well as young researchers starting out in a career in academia. In addition, every paper was assigned to a topic area. ICDE 2016 was structured into 15 topic areas handled by individual area chairs. With topics ranging from Data Analytics and Big Data to Systems with Modern Hardware and In-Memory Database Architectures, ICDE 2016 provided a premier forum for the exchange and dissemination of data management research results among researchers, users, practitioners, and developers. In addition to research papers, the conference also offered an Industrial and Applications track with 12 papers accepted out of 44 submissions as well as a demo track with 68 submissions and 23 accepted demos.

Out of the 99 papers presented at the conference, six outstanding papers were selected and the authors were offered the opportunity to provide an extended version to be published in this special section of the IEEE Transactions on Knowledge and Data Engineering. After an additional—and again—rigorous review process, all papers are now substantially extended in comparison to their conference version. Those six papers cover a wide range of topics and very nicely reflect the breadth of research work in the context of data management.

The paper “l-Injection: Toward Effective Collaborative Filtering Using Uninteresting Items” by Jongwuk Lee, Sang-Wook Kim, Won-Seok Hwang, Juan Parc, Youngnam Lee, and Dongwon Lee addresses the challenge of sparsity in recommender systems. By carefully injecting low values to a selected set of unrated user-item pairs in a user-item matrix, they demonstrate that top-N recommendation accuracies of various collaborative filtering (CF) techniques can be significantly and consistently improved. Through comprehensive experiments with three real-life datasets (e.g., Movielens, Ciao, and Watcha), they demonstrate that their solution consistently and universally enhances the accuracies of existing CF algorithms (e.g., item-based CF, SVD-based CF, and SVD + +) by 2.5 to 5 times on average.

The paper “Platform-Independent Robust Query Processing” by Srinivas Karthik, Jayant R. Haritsa, Sreyash Kenkre, Vinayaka Pandit, and Lohit Krishnan tackles the classical selectivity estimation problem for OLAP queries in relational databases, by focusing on a few problems that a new approach called PlanBouquet had. They introduce SpillBound, a new query processing algorithm which retains the core strength of PlanBouquet’s discovery process and provides better performance. They also prove that SpillBound is within an \( O(D) \) factor of the best possible deterministic algorithm in its class, where \( D \) is the number of error-prone predicates in the query.

The paper “Passive and Partially Active Fault Tolerance for Massively Parallel Stream Processing Engines” by Li Su and Yongluan Zhou studies fault-tolerance techniques for stream processing engines which can be categorized into passive and active approaches, and the problems they have in Massively Parallel Stream Processing Engines (MPSPE). They propose a new fault-tolerance framework, which is Passive and Partially Active (PPA); in such a scheme, the passive approach is applied to all tasks while only a selected set of tasks will be actively replicated. They propose effective and efficient algorithms to optimize a partially active replication plan to maximize the quality of tentative outputs. They also discuss the implementation of PPA on top of Storm, and extensive experiments using both real and synthetic datasets.

The paper “Interactive Data Exploration with Smart Drill-Down” by Manas Joglekar, Hector Garcia-Molina, and Aditya Parameswaran considers smart drill-down, an operator for interactively exploring a relational table to discover and summarize interesting groups of tuples, where “interesting” is characterized via a specific rule. For instance, the rule \((a, b, *, 1000)\) may indicate that there are 1000 tuples with value \(a\) in the first column and \(b\) in the second column (and any value in the third column). The paper shows that the underlying optimization problems are NP-Hard and describe an algorithm for finding the approximately optimal list of rules to display when the user uses a smart drill-down,
and a dynamic sampling scheme for efficiently interacting with large tables. The paper also reports on experiments on real datasets to demonstrate the usefulness of smart drill-down and study the performance of their algorithms.

The paper “Inferring Higher-Order Structure Statistics of Large Networks from Sampled Edges” by Pinghui Wang, Yiyian Qi, John C.S. Lui, Don Towsley, Junzhou Zhao, and Jing Tao looks at the problem of exploring locally connected subgraphs (also known as motifs or graphlets) of complex networks. Previous work made the strong assumption that the graph topology of interest is known in advance. In practice, what is available is only a snapshot of the graph, i.e., a subgraph of the graph. Crawling methods such as breadth first sampling can be used to generate the snapshot, but these methods fail to sample a streaming graph represented as a high-speed stream of edges. The paper proposes a random edge sampling method to collect edges and generate a sampled graph, which they call a RESampled graph. As a RESampled graph’s motif statistics may be quite different from those of the original graph, the authors propose the Minfer framework, which takes the given RESampled graph and accurately infers the underlying graph’s motif statistics.

The paper “I/O Efficient Core Graph Decomposition: Application to Degeneracy Ordering” by Dong Wen, Lu Qin, Ying Zhang, Xuemin Lin, and Jeffrey Xu Yu looks at the problem of core decomposition which is a fundamental graph problem with a large number of applications. Most existing approaches for core decomposition assume that the graph is kept in memory of a machine while many real-world graphs are too big to reside in memory. The paper studies I/O efficient core decomposition following a semi-external model, which only allows node information to be loaded in memory. The authors propose a semi-external algorithm and an optimized algorithm for I/O efficient core decomposition, handling dynamic graphs as well. In addition, they further propose an I/O efficient semi-external algorithm for degeneracy ordering, which is an important graph problem that is highly related to core decomposition.

ACKNOWLEDGMENTS

The Guest Editors would like to thank all of the authors for meticulously revising and extending their papers to address the comments raised by the conference and journal reviewers. Their appreciation also goes to the program committee members and especially all area chairs for their effort in making the research program of ICDE 2016 a great success. They also take this opportunity to thank all of the parties involved in making the conference a tremendous success. Last, but not least, we sincerely hope that you will enjoy this special section; we are convinced that this special section provides a very nice collection of high-quality research papers in the field of data engineering.

Meichun Hsu
Alfons Kemper
Timos Sellis

Guest Editors and ICDE 2016 Program Committee Chairs

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.