

Human-in-the-loop Exploration of Composite Items

Senjuti Basu Roy

New Jersey Institute of Technology

Newark, NJ, USA

senjuti.basuroy@njit.edu

ABSTRACT

Human-in-the-loop data exploration is seeing a renewed interest in our community. With the rise of big data analytics, this area is growing to encompass not only approaches and algorithms to find the next best data items to explore but also interactivity, i.e. accounting for feedback from the data scientist during the exploration. Interactivity is essential to account for evolving needs during the exploration and also customize the discovery process. In this tutorial, we focus on exploration of Composite Items (CIs) that require repeated interaction with human users.

CIs address complex information needs and are prevalent in online shopping where products are bundled together to provide discounts, in travel itinerary recommendation where points of interest in a city are combined into a single travel package, and task assignment in crowdsourcing where personalized micro-tasks are composed and recommended to workers. CI formation is usually expressed as a constrained optimization problem. For instance, in online shopping, package retrieval can retrieve the cheapest smartphones (optimization objective) with compatible accessories (constraints). Similarly, a city tour must be the most popular and conform to a total time and cost budget. A data scientist interested in exploring a variety of CIs has to repeatedly reformulate optimization problems with new constraints and objectives. In this tutorial, we investigate the applicability of interactive data exploration approaches to CI formation.

The tutorial will have the following parts:

- **15 minutes** We will first review CI applications and shapes (15mn) that are applicable in different domains. This part will gather different examples and attempt to unify them.
- **60 minutes** We then discuss three big research questions : (i) existing algorithms for CI formation, (ii) human-in-the-loop CIs, and (iii) optimization opportunities.
- **15 minutes** We conclude with ongoing and future research directions.

The tutorial targets theoreticians and practitioners interested in the development of data science applications. It should be of particular interest to database researchers, applied machine learners, as well as data scientists in industrial research settings who want to learn about how different domains, such as product recommendation, scientific simulation, or team formation in the social sciences and crowdsourcing, have been developing their “siloed” definitions of CIs. The research direction presented in the tutorial will be helpful

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

COMAD 2019, January 3-5, Kolkata, India

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-3-89318-078-3.

DOI: 10.1145/3297001.3297065

to converge these domain specific ideas and creating an overarching generic framework. Tutorial attendees are expected to have basic knowledge in algorithms and data management. Knowledge in constrained optimization is not necessary.

The proposed tutorial is timely. It brings together several related efforts and addresses unsolved questions in the emerging area of human-in-the-loop exploration of complex information needs. The tutorial is relevant to the general area of data science and more specifically to Scalable Analytics, Data Mining, Clustering and Knowledge Discovery, Indexing, Query Processing and Optimization, and Crowdsourcing. The technical topics covered are constrained optimization, ranking semantics, clustering, algorithms, and empirical evaluations.

BIOGRAPHY

Senjuti Basu Roy, *senjuti.basuroy@njit.edu*, is an Assistant Professor at NJIT. She has received her M.S and PhD degree in Computer Science from the University of Texas at Arlington and bachelors in technology (B.Tech) degree from University of Calcutta. Her industrial experience includes working at Microsoft Research and IBM Research.

Senjuti’s broader research interests lie in the area of data and content management of web and structured data with a focus on exploration, analytics, and algorithms. In recent years, her research has focused on designing principled algorithms and systems that require man-machine collaboration. Senjuti has presented three tutorials on the computational challenges related to man-machine systems in Very Large Database Conference, International World Wide Web Conference, and International Conference on Extending Database Technology. She was the PC Co-chair of SIGMOD 2018 mentorship track, and the PC co-chair of VLDB 2018 PhD Workshop program. Senjuti was a co-organizer of ExploreDB 2016 (co-located with SIGMOD 2016) and the IEEE Workshop on Human-in-the-loop Methods and Human Machine Collaboration in BigData (IEEE HMDData 2017, 2018) (co-located with IEEE Big data). She has organized a NSF workshop on converging human and technological perspectives in crowdsourcing research and will be organizing a Shonan Meeting on the topic of “Human-in-the-loop Big Data and AI: Connecting Theories and Practices for a Better Future of Work”.

Senjuti has published more than 55 prestigious research papers in top-tier international conferences and journals. She has published seminal papers on composite item retrieval and data exploration.

ACKNOWLEDGMENTS

The work is supported by the National Science Foundation under Grant No.: 1814595 and Office of Naval Research under grant name: Human-AI Symbiosis for Agile Planning.