Explore-By-Example: A New Database Service for Interactive Data Exploration

Yanlei Diao
University of Massachusetts Amherst
yanlei@cs.umass.edu

Abstract

Traditional DBMSs are suited for applications in which the structure, meaning and contents of the database, as well as the questions (queries) to be asked, are all well-understood. However, this is no longer true when the volume and diversity of data grow at an unprecedented rate, while the user ability to comprehend data remains (as limited) as before. To address the increasing disparity in the "big data - same humans" problem, our project explores a new approach of system-aided exploration of a big data space and automatic learning of the user interest in order to retrieve all objects that match the user interest – we call this new service "interactive data exploration", which complements the traditional querying interface of a database system.

In this talk, I introduce a new framework for interactive data exploration, called "Explore-by-Example", which iteratively seeks user relevance feedback on database samples and uses such feedback to finally predict a query that retrieves all objects of interest to the user. The goal is to make such exploration converge fast to the true user interest model, while minimizing the user labeling effort and providing interactive performance in each iteration. I discuss a range of techniques and optimizations to do so for linear patterns and complex non-linear patterns. Our user study indicates that our approach can significantly reduce the user effort and the total exploration time, compared with the common practice of manual exploration. I finally conclude the talk by pointing out a host of new challenges, ranging from application of active learning theory, to database optimizations, to visualization.

Short Bio

Yanlei Diao is an Associate Professor of Computer Science at the University of Massachusetts Amherst. Her research interests are in information architectures and data management systems, with a focus on big data analytics, scientific analytics, data streams, uncertain data management, and RFID and sensor data management. She received her PhD in Computer Science from the University of California, Berkeley in 2005, her M.S. in Computer Science from the Hong Kong University of Science and Technology in 2000, and her B.S. in Computer Science from Fudan University in 1998.

Yanlei Diao was a recipient of the 2013 CRA-W Borg Early Career Award (one female computer scientist selected each year), IBM Scalable Innovation Faculty Award, and NSF Career Award, and she was a finalist of the Microsoft Research New Faculty Award. She spoke at the Distinguished Faculty Lecture Series at the University of Texas at Austin. Her PhD dissertation "Query Processing for Large-Scale XML Message Brokering" won the 2006 ACM SIGMOD Dissertation Award Honorable Mention. She is currently Editor-in-Chief of the ACM SIGMOD Record, Associate Editor of ACM TODS, Associate Editor of VLDB 2016, and member of the SIGMOD Executive Committee and SIGMOD Software Systems Award Committee. In the past, she has served as Associate Editor of PVLDB, Area Chair of SIGMOD, organizing committee member of SIGMOD, CIDR, DMSN, and the New England Database Summit, as well as on the program committees of many international conferences and workshops.

ExploreDB'15, May 31 - June 04 2015, Melbourne, VIC, Australia.
ACM 978-1-4503-3740-3/15/05.
http://dx.doi.org/10.1145/2795218.2795226.