Future Directions of Humans in Big Data Research

Summary of the 1st Workshop on Human-Centered Big Data Research

Celeste Lyn Paul
Department of Defense
Fort Meade, MD, United States
clpaul@tycho.ncsc.mil

Chris Argenta
Applied Research Associates
Raleigh, NC, United States
cargenta@ara.com

William Elm
Resilient Cognitive Solutions
Pittsburgh, PA, United States
welm@resilientcognitivesolutions.com

Alex Endert
Georgia Institute of Technology
Atlanta, GA, United States
endert@gatech.edu

***Abstract*—The goal of the *1st Workshop on Human-Centered Big Data Research* was to explore the multi-disciplinary challenges of researching humans in Big Data environments. This paper summarizes the outcomes of the workshop and aims to define potential future work in this area.**

Keywords—Big data; human-centered computing; workshop.

I. INTRODUCTION

Big Data are data with high *volume* (number of records), *velocity* (speed of acquisition), and/or *variety* (complexity). The traditional definition of Big Data used to be described as more data than could fit into memory for processing or could be stored. However, the modern interpretation of Big Data is that size is relative to the type of problem you are working on or the context you are working in. What "big" data is for a small business is different than what "big" data is for a genetics researcher.

Most discussion (and research) about Big Data has focused on the technical aspects of Big Data: storing more, processing faster, and improving algorithm efficiency. However, little of this focus has investigated the human factors of Big Data. Yet, humans are the ones who consume and find value in Big Data.

There is significant related work to humans and Big Data that pre-dates the modern sense of Big Data, such as information overload [3], sensemaking of large and complex data [5], and visual analytics to support analysis of big data through highly visual interfaces [6]. Principles from this research are applied to Big Data systems; but, no new research has emerged since the modern Big Data age.

While Big Data is often thought of as technology, it is also a very multidisciplinary problem. As in human-centered computing, research in not only technology but also psychology, design, and knowledge of application domains are necessary as well. What we need are new research questions that contribute to new phases of human and Big Data research and a research framework on how to conduct or share this research.

II. WORKSHOP ON HUMAN-CENTERED BIG DATA RESEARCH

The 1st Workshop on Human-Centered Big Data Research was held at the Laboratory for Analytic Sciences at North Carolina State University on April 1-3, 2014. The purpose of the workshop was to identify human-centered Big Data research as an interesting and important research area, be a forum for researchers in different disciplines to share their research and provide a valuable opportunity to discuss the challenges of this research area. The goal of the workshop was threefold. First, the workshop was the first step toward creating a research community focused on human-centered Big Data research. Second, it was an opportunity to share information across a small community of researchers. Finally, the workshop was an opportunity for these researchers to work together to begin defining the research space by developing a common language, identifying research questions, and proposing research methodologies.

There was very good response to the call for papers. In total, 72 submissions were received and 10 were accepted with an acceptance rate of 14%. Each submission was reviewed by at least three of the five committee members. Authors represented a wide range of disciplines including business, psychology, sociology, and computer science. Two additional speakers relevant to the workshop topic were also invited to present as well as a keynote from the director of the Laboratory for Analytic Sciences.

The workshop hosted approximately 40 participants. Half of the participants were accepted paper speakers and invited participants from universities, laboratories, and commercial research companies. The remainder of the participants came from agencies in the U.S. government who had an interest in research related to the analytic challenges of Big Data. The workshop took place over the course of two and a half days. On days one and two, the morning sessions were dedicated to technical talks from invited speakers. The afternoon sessions were dedicated to one of three breakout sessions. On day three, the morning session was dedicated to presentations of the results of the three breakout sessions and a final group discussion of take aways from the workshop.

978-1-4799-5666-1/14/$31.00 ©2014 IEEE

III. Workshop Results

A. Research Presentations & Group Discussions

There were a variety of topics presented at the workshop including: attention and cognitive work, cognitive and mental modeling, tacit knowledge in specialized domains, visualization for situation awareness, and complex problems such as business intelligence and computer network defense.

Although the research was very diverse, all of the presentations described the challenges of human-centered research in Big Data environments. One of the common themes across many of the presentations was the need for experts to help make sense of Big Data. These experts are critical for understanding work processes, evaluating user models, and developing new analytics. However, experts are also notoriously difficult to involve in research. Experts tend to be very busy and unavailable for researchers to engage with.

B. Breakout Sessions

Participants separated into one of three breakout sessions in the afternoons of day one and day two. The purpose of these breakout sessions were to discuss the challenges and research questions associated with one of three topic areas in Big Data environments: sensemaking, modeling, and methodology.

1) Sensemaking in Big Data Environments

The sensemaking breakout session focused on defining and discussing ways to support research of user analysis processes in Big Data environments. These discussions resulted in a proposed framework for assessing sensemaking in Big Data environments. The four dimensions of the framework are:

- **Human:** Demographics, personality, teaming…
- **Tasks:** Find *x*, relationship of *x* and *y*, similarity to *x*…
- **Tools:** Analytics, visualization, applications…
- **Big Data:** Volume, velocity, variety…

The framework is proposed as a community tool for focusing and communicating human-centered Big Data research. The framework will support systematically exploring and evaluating sensemaking in a Big Data environment. An essential element of this framework is to enable researchers to communicate how their work contributes to the larger body of human-centered Big Data research.

2) Modeling in Big Data Environments

The modeling breakout session focused on defining and identifying major challenges associated with modeling in human-centered Big Data environments. The group identified two types of models important in Big Data environments: the user's model of the world and the system's model of the world.

Use of modeling poses three challenges for research:

- Bi-directional representation of user and system models
- User's trust and understanding of the system model
- Impact of user and system models on analytic culture

The definition of user and system models helps researchers in different disciplines more easily communicate their work to others. The modeling challenges posed by the group provide interesting research goals to pursue in future work.

3) Methodology in Big Data Environments

The methodology breakout session focused on discussing and developing ways in which users can be studied in Big Data environments. These discussions resulted in a proposed framework for selecting and assessing methods in Big Data environments. The three dimensions of the framework are:

- **Organizations:** policies, procedures, pressures…
- **Tools and Data:** functionality, V's, training…
- **Users:** mental models, attention, internal framing…

The goal of the framework is to help researchers consider the holistic Big Data environment. Research methodologies such as ethnography excel at understanding the Organization and User while instrumentation provides a non-invasive way to understand the use of Tools and Data.

IV. Next Steps

Discussion of the challenges of humans in Big Data research and future directions of research in this area continue. Many of the academic, laboratory, and industry participants found it valuable to talk to government participants and learn more about their Big Data problems. Likewise, government participants found it valuable to talk to researchers first-hand about their research and how it may be applicable to their missions. Proceedings of the workshop are published in the ACM International Conference Proceedings Series (ICPS) and available in the ACM Digital Library [4]. Videos of the talks are hosted on the Laboratory for Analytic Sciences website [2]. Additionally, an open mailing list supports continued discussion in human-centered Big Data research [1].

Acknowledgments

Thanks to the Laboratory for Analytic Sciences and the North Carolina State University for hosting the workshop, the authors of the presentations, and government attendees.

References

[1] HCBDR mailing list. hcbdr@googlegroups.com.

[2] Laboratory for Analytic Sciences. http://ncsu-las.org/.

[3] Levy, D.M. To Grow in Wisdom: Vannevar Bush, Information Overload, and the Life of Leisure. Proceedings of the Joint Conference on Digital Libraries, 2005, 281-286.

[4] Paul, C.L., Harris, J.D., Keiser, J., Moore, K. and Allen, F. (eds.) *Proceedings of the 1st Workshop on Human-Centered Big Data Research*. Laboratory for Analytic Sciences, North Carolina State University, April 1-3, 2014. ACM International Conference Proceeding Series, 2014.

[5] Pirolli, P. and Card, S. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis*, McClean, VA, May 2-6, 2005.

[6] Thomas, J.J. and Cook, K. (eds.) *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.