Chapter 76

Beyond Micro-Tasks: Research Opportunities in Observational Crowdsourcing

Roman Lukyanenko
University of Saskatchewan, Canada

Jeffrey Parsons
Memorial University of Newfoundland, Canada

ABSTRACT

The emergence of crowdsourcing as an important mode of information production has attracted increasing research attention. In this article, the authors review crowdsourcing research in the data management field. Most research in this domain can be termed task-based, focusing on micro-tasks that exploit scale and redundancy in crowds. The authors’ review points to another important type of crowdsourcing – which they term observational – that can expand the scope of extant crowdsourcing data management research. Observational crowdsourcing consists of projects that harness human sensory ability to support long-term data acquisition. The authors consider the challenges in this domain, review approaches to data management for crowdsourcing, and suggest directions for future research that bridges the gaps between the two research streams.

INTRODUCTION

Recent years have seen a major shift in knowledge production via crowdsourcing, wherein increasingly work is being done by distributed members of the general public (the crowd), rather than employees or traditional subsidiaries. Crowdsourcing promises to dramatically expand organizational computing power and “sensor” networks, making it possible to engage ordinary people in large-scale data collection (Brabham, 2013; Doan, Ramakrishnan, & Halevy, 2011; Franklin, Kossmann, Kraska, Ramesh, & Xin, 2011; Garcia-Molina, Joglekar, Marcus, Parameswaran, & Verroios, 2016; Li, Wang, Zheng, & Franklin, 2016).
Applications of crowdsourcing are rapidly expanding and power such diverse activities as corporate product development, marketing, public policy, scientific research, graphic design, software development, and writing and editing. Crowdsourcing is increasingly tasked with tackling difficult societal and technological challenges, such as climate change (Theobald et al., 2015), natural disasters (Brabham, 2013) and commonsense reasoning in artificial intelligence (Davis & Marcus, 2015).

Organizations integrate crowdsourcing into internal decision making and operations. Fortune 500 companies maintain digital platforms to monitor what potential customers are saying and understand customer reactions to products and services. They also use consumer feedback to design better products and monitor market changes (Abbasi, Chen, & Salem, 2008; Barwise & Meehan, 2010; Brynjolfsson & McAfee, 2014; Delort, Arunasalam, & Paris, 2011).

Crowdsourcing turns problem-solving capacity and data into commodities, making them available on-demand. For example, many municipalities in the United States now subscribe to CitySourced.com, which harnesses citizens’ reports of crime, graffiti, potholes, broken street lights, and other civic issues, to better support infrastructure management. In a more general setting, Amazon’s Mechanical Turk (mturk.com), CrowdFlower.com, and Clickworker.com maintain pools of “crowdworkers” that companies hire on-demand to perform small problem-solving tasks.

There is also a proliferation of platforms for automatically generating data collection forms that can be easily configured and rapidly launched on a large scale. Projects such as EpiCollect.net, SciStarter.com, or SmartCitizen.me make crowdsourcing possible for organizations and even individuals, requiring little technical expertise and infrastructure. Crowd-powered extensions of word processors, such as Soylent, enlist crowds for document writing and editing (Bernstein et al., 2015). In addition to becoming a mainstream commercial service, crowdsourcing has become a major resource for scientific research (Goodman & Paolacci, 2017).

Crowdsourcing presents several data management challenges. Unlike traditional data collection in organizations, in crowdsourcing there are typically weaker constraints on who can participate. This creates the challenge of managing data produced by often anonymous users with varying levels of domain expertise or motivation (Lukyanenko, Parsons, Wiersma, Sieber, & Maddah, 2016). Furthermore, in many projects participation is voluntary, making it difficult to engage users in eliciting information requirements (e.g., to guide database design) or in improving the quality of existing data (e.g., to clarify a particular data entry, or request additional information) (Chen, Xu, & Whinston, 2011). These challenges offer exciting opportunities for data management researchers to design innovative solutions.

In this paper, we survey current research on data management in crowdsourcing. We argue that two major, largely disconnected, streams of crowdsourcing research exist: (1) studies that investigate uses for crowds as links in a larger technological chain and tend to focus on small, granular and well-defined micro-tasks (task-based crowdsourcing); (2) studies that explore the potential of crowds as “sensors” in the environment that can be leveraged in organizational decision making and innovation (observational crowdsourcing). Until now, the data management community has studied primarily the first kind of crowdsourcing, resulting in a serious gap in resolving challenges associated with the latter type. With this in mind, we analyze the challenges and opportunities associated with observational crowdsourcing environments and suggest directions for future database research in this context.