Human Factors in Crowdsourcing

Sihem Amer-Yahia[‡], Senjuti Basu Roy[††]
[‡]University Grenoble Alps, CNRS, France, [††]New Jersey Institute of Technology, USA
[‡]sihem.amer-yahia@imag.fr,[††]senjutib@njit.edu

1. ABSTRACT AND GOAL

Today, crowdsourcing is used to "taskify" any job ranging from simple receipt transcription to collaborative editing, fan-subbing, citizen science, and citizen journalism. The crowd is typically volatile, its arrival and departure asynchronous, and its levels of attention and accuracy diverse. Tasks vary in complexity and may necessitate the participation of workers with varying degrees of expertise. Sometimes, workers need to collaborate explicitly and build on each other's contributions to complete a single task. For example, in disaster reporting, CrowdMap allows geographically closed people with diverse and complementary skills, to work together to report details about the course of a typhoon or the aftermath of an earthquake.

This uber-ization of human labor requires the understanding of workers motivation in completing a task, their ability to work together in collaborative tasks, as well as, helping workers find relevant tasks. For over 40 years, organization studies have thoroughly examined human factors that affect workers in physical workplaces. More recently, computer scientists have developed algorithms that verify and leverage those findings in a virtual marketplace, in this case, a crowdsourcing platform.

The goal of this tutorial is to review those two areas and discuss how their combination may improve workers' experience, task throughput and outcome quality for both microtasks and collaborative tasks. We will start with a coverage of motivation theory, team formation, and learning worker profiles. We will then address open research questions that result from this review.

2. TOPIC AND DESCRIPTION

The proposed tutorial brings together two research areas in order to understand the new challenges and opportunities in making crowdsourcing more effective. It will start with a review of work in organization studies that is founded in motivation theory, its formalization of human factors, and the findings that have been observed in physical workplaces.

Proceedings of the VLDB Endowment, Vol. 9, No. 13

The second part of this tutorial will focus on reviewing work by computer scientists on quantifying human factors and on studying their impact on task assignment to workers and on task completion. The last part will discuss future directions that are raised by bringing together expertise accumulated by social scientists and models and algorithms developed by computer scientists.

Existing work in organization studies has focused on understanding and quantifying human factors that influence the ability of an individual to perform a task, or a set of tasks, alone, or in collaboration with others, for over 40 years. In that context, workers' motivation is modeled as a combination of *intrinsic* and *extrinsic* factors. Extrinsic motivation is an instrument for achieving a certain desired outcome (e.g. making money or improving ones social image). Intrinsic motivation exists if an individual works for fulfillment generated by the activity (e.g. working just for fun or achieving a humanitarian goal). Examples of intrinsic factors are task significance (the degree to which the task has an influence over other people), and task feedback (the extent to which precise information about the effectiveness of performance is conveyed to the worker). It has been shown that such human factors highly depend on the nature of tasks worker undertake and directly affect their motivation during task completion. We will describe how work motivation theory has modeled and validated those factors in physical workplaces. We will additionally review factors that affect the ability of workers to interact with other workers, such as affinity and critical mass.

In computer science, the focus has been on developing models and algorithms for task assignment and for task completion. On the task assignment front, there has been substantial research on team formation which is appropriate for collaborative tasks. Team formation is prevalent in social networks analysis and in group recommendations. We will review work that shows how popular crowdsourcing applications such as fan-subbing or document editing benefit from team formation where the set of workers with the best human factors is assigned to tasks. We will also review the complexity of algorithms used to assign tasks to workers and their behaviour in practice. On the task completion front, existing work has mainly focused on observing workers while they complete tasks, learning their humans factors (primarily, their skills), and devising strategies to improve workers' experience and overall outcome quality. A number of efforts have been on verifying if the correlations between human factors, such as doing a task for fun and enjoyment or for advancing human capital, hold in a crowdsourcing

marketplace. In order to keep workers interested in task completion, existing work proposed incentivizing workers for long-lasting tasks or entertaining workers during task completion.

Despite recent advances in computer science on crowdsourcing models and algorithms, there is still a gap to fill in how to best leverage findings from organization studies and marketplace design on a crowdsourcing platform. This tutorial will be dedicated to bridging that gap.

3. DURATION AND SESSIONS

The turorial will last **90mn**. It will review relevant work in organization studies and in computer science and will decidate **40mn** to new opportunities raised by bringing together those two areas.

3.1 PART 1: Human Factors at Work

We will dedicate **20mn** to defining and reviewing human factors at work. This part contains a condensed overview of formalizations and studies of humans in physical and virtual workplaces and their impact on workers' performance, satisfaction and job outcome. This part relies on theoretical foundations in motivation theory and job redesign. We will also review work that verifies findings related to task completion in physical workplaces in a crowdsourcing marketplace.

3.2 PART 2: Algorithms for Task Assignment and Task Completion

We will dedicate **20mn** to reviewing models and algorithms. This part covers devising appropriate task assignment strategies that rely on human factors, for both microtasks and collaborative tasks.

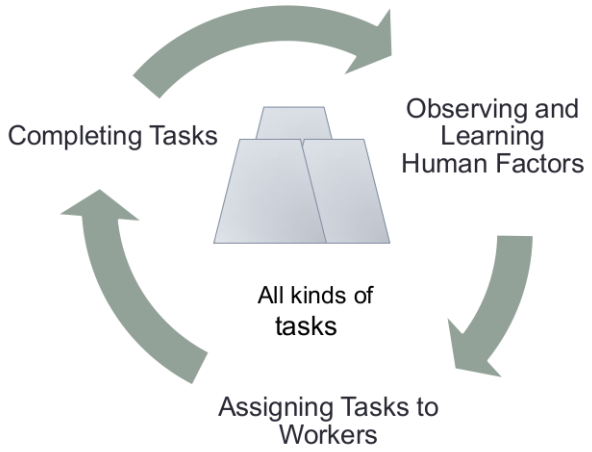

Figure 1: Crowdsourcing Processes

3.3 PART 3: Future Directions

We will dedicate **40mn** to new challenges. This part discusses the new research opportunities raised by integrating human factors described in PART 1 into algorithms described in PART 2. We will discuss how additional dimensions in modeling workers' motivation and other factors can enrich crowdsourcing processes as illustrated in Figure 1. We will also discuss result validation.

4. AUDIENCE AND PREREQUISITE

The tutorial targets theoreticians and practitioners interested in the development of data-centric applications in crowdsourcing. It should be of particular interest to an audience who wants to learn about how the social sciences have been modeling motivation at work and how those models should help design and verify different algorithms task assignment, worker profile learning and task completion.

Tutorial attendees are expected to have basic knowledge in algorithms and data management. Knowledge in constrained optimization is not necessary.

5. RELEVANCE

The proposed tutorial is timely as it addresses unsolved questions in the emerging area of crowdsourcing. The tutorial is relevant to data management and the web and more specifically, to Big Data Processing and Transformation, Data Mining, Clustering and Knowledge Discovery, Large-Scale Analytics, Indexing, Query Processing and Optimization, Social Networks and Analysis, Graph Databases, Information Retrieval, and Modeling, Mining and Querying User Generated Content. The technical topics covered are work motivation theory, constrained optimization, hardness results and algorithms, and empirical evaluations. Further details are described in the format and outline. The authors published seminal papers in crowdsourcing and team formation and group recommendation. Their names are highlighted in bold in the list of references.

6. PREVIOUS EDITIONS

This tutorial is under consideration at SIGMOD.

A first version of this tutorial focusing solely on reviewing Computer Science research in crowdsourcing, was presented at WWW in 2015. That version was focused on Part 2 only. It did not include Part 1 and therefore did not study Part 2 in conjunction with Part 1.

7. OTHER TOPICALLY RELATED TUTORIALS

We reviewed the last 5 years of PVLDB, SIGMOD, ICDE, EDBT, WWW, ICWSM, WSDM and found the following related tutorials. the closest in spirit is the last one in the list that examines the population of workers and their strategic behavior (from a game theory viewpoint) on crowdsourcing platforms.

1. Truth Discovery and Crowdsourcing Aggregation: A Unified Perspective Jing Gao, Qi Li, Bo Zhao, Wei Fan, Jiawei Han. PVLDB 2015.

2. Data-driven Crowdsourcing: Management, Mining and Applications. Lei Chen, Dongwon Lee, Tova Milo. ICDE 2015.

3. From Complex Object Exploration to Complex Crowdsourcing. Sihem Amer-Yahia, Senjuti Basu Roy. WWW 2015.

4. Social Spam, Campaigns, Misinformation and Crowdturfing.Organizers: Kyumin Lee, James Caverlee and Calton Pu. WWW 2014.

5. How to use Mechanical Turk for Behavioral Research. Winter Mason, Siddharth Suri. ICWSM 2011.

8. BIOS

Sihem Amer-Yahia is DR1 CNRS at LIG in Grenoble where she leads the SLIDE team. Before joining CNRS, she was Principal Scientist at the Qatar Computing Research Institute, Senior Scientist at Yahoo! Research and at&t Labs. Sihem has served on the SIGMOD Executive Board and on the PVLDB and the EDBT Endowments. She is the Editor-in-Chief of the VLDB Journal for Europe and Africa and is on the editorial boards of TODS and the Information System Journal. She is currently serving as area chair for SIGMOD 2016 and is co-chairing the PVLDB 2016 Workshops. Sihem received her Ph.D. in Computer Science from Paris-Orsay and INRIA in 1999.

Senjuti Basu Roy is an Assistant Professor at the Institute of Technology at the University of Washington Tacoma. Prior to joining UW in January 2012, she was a postdoctoral fellow at DIMACS at Rutgers University. Senjuti received her Ph.D. in Computer Science from UT Arlington in May 2011. Her research interests lie in the data and content management of web and structured data with a focus on exploration, analytics, and algorithms. Her research has been published in premier database and IR conferences and journals. Her past industrial experience includes working at Microsoft Research and IBM Research. She has co-chaired KDD Cup 2013, ExploreDB 2016, and serves as a guest editor of JMLR special issue on KDD Cup 2013.

9. REFERENCES

We have reviewed several papers in relation with the topic of our tutorial and selected the most relevant papers. We present those papers grouped by topic and rank them in alphabetical order. We present papers of which one of us is a co-author at the end of each group.

9.1 In Crowdsourcing

1. Jacob Abernethy, Rafael M. Frongillo: A Collaborative Mechanism for Crowdsourcing Prediction Problems. NIPS 2011: 2600-2608.

2. Paul Andr, Haoqi Zhang, Juho Kim, Lydia B. Chilton, Steven P. Dow, Robert C. Miller: Community Clustering: Leveraging an Academic Crowd to Form Coherent Conference Sessions. HCOMP 2013.

3. Lydia B. Chilton, Juho Kim, Paul Andr, Felicia Cordeiro, James A. Landay, Daniel S. Weld, Steven P. Dow, Robert C. Miller, Haoqi Zhang: Frenzy: collaborative data organization for creating conference sessions. CHI 2014: 1255-1264.

4. Juho Kim, Haoqi Zhang, Paul Andr, Lydia B. Chilton, Anant Bhardwaj, David R. Karger, Steven P. Dow, Robert C. Miller: Cobi: Community-Informed Conference Scheduling. HCOMP 2013.

5. Edith Law, Haoqi Zhang: Towards Large-Scale Collaborative Planning: Answering High-Level Search Queries Using Human Computation. AAAI 2011.

6. Robert C. Miller, Haoqi Zhang, Eric Gilbert, Elizabeth Gerber: Pair research: matching people for collaboration, learning, and productivity. CSCW 2014: 1043-1048.

7. Haoqi Zhang, Andrs Monroy-Hernndez, Aaron D. Shaw, Sean A. Munson, Elizabeth M. Gerber, Benjamin Mako Hill, Peter Kinnaird, Shelly Diane Farnham, Patrick Minder: WeDo: End-To-End Computer Supported Collective Action. ICWSM 2014.

8. Atsuyuki Morishima, **Sihem Amer-Yahia**, **Senjuti Basu Roy**: Crowd4U: An Initiative for Constructing an Open Academic Crowdsourcing Network. HCOMP 2014.

9. Habibur Rahman, **Senjuti Basu Roy**, Saravanan Thirumuruganathan, **Sihem Amer-Yahia**, Gautam Das: "The Whole Is Greater Than the Sum of Its Parts": Optimization in Collaborative Crowdsourcing. CoRR abs/1502.05106 (2015).

10. Habibur Rahman, **Senjuti Basu Roy**, Saravanan Thirumuruganathan, **Sihem Amer-Yahia**, Gautam Das: Task Assignment Optimization in Collaborative Crowdsourcing. ICDM 2015.

11. Habibur Rahman, Saravanan Thirumuruganathan, **Senjuti Basu Roy**, **Sihem Amer-Yahia**, Gautam Das: Worker Skill Estimation in Team-Based Tasks. PVLDB 8(11): 1142-1153 (2015).

12. **Senjuti Basu Roy**, **Sihem Amer-Yahia**, Lucas Joppa: ECCO- A Framework for Ecological Data Collection and Management Involving Human Workers. EDBT 2015: 677-682.

13. **Senjuti Basu Roy**, Ioanna Lykourentzou, Saravanan Thirumuruganathan, **Sihem Amer-Yahia**, Gautam Das: Task assignment optimization in knowledge-intensive crowdsourcing. VLDB J. 24(4): 467-491 (2015).

14. Beatrice Valeri, Shady Elbassuoni, **Sihem Amer-Yahia**: Acquiring Reliable Ratings from the Crowd. HCOMP 2015: 40-41.

15. Kosetsu Ikeda, Atsuyuki Morishima, Habibur Rahman, **Senjuti Basu Roy**, Saravanan Thirumuruganathan, **Sihem Amer-Yahia**, Gautam Das. Collaborative Crowdsourcing with Crowd4U. VLDB 2016 (demonstration)

9.2 In Quantifying and Leveraging Motivation at Work

1. B. B. Bederson and A. J. Quinn. Web workers unite! addressing challenges of online laborers. CHI 2011, Extended Abstracts Volume, Vancouver, BC, Canada, May 7-12, 2011, 97-106, 2011.

2. D. Chandler and A. Kapelner. Breaking monotony with meaning: Motivation in crowdsourcing markets. Journal of Economic Behavior & Organization, 90:123-133, 2013.

3. P. Dai, J. M. Rzeszotarski, P. Paritosh, and E. H. Chi. And now for something completely different: Improving crowdsourcing workflows with micro-diversions. CSCW, 628-638, 2015.

4. Yihan Gao, Aditya G. Parameswaran: Finish Them!: Pricing Algorithms for Human Computation. PVLDB 7(14): 1965-1976 (2014).

5. J. Hackman and G R Oldham. Motivation through the design of work: Test of a theory. Organizational behavior and human performance, 16(2):250279, 1976.

6. G. Hertel and G. Hertel, Synergetic effects in working teams. Journal of Managerial Psychology, 2011.

7. P. G. Ipeirotis. Analyzing the amazon mechanical turk marketplace. ACM Crossroads, 17(2):16-21, 2010.

8. N. Kaufmann, T. Schulze, and D. Veit. More than fun and money. worker motivation in crowdsourcing A study on mechanical turk. AMCIS, 2011

9. A. Kittur, J. V. Nickerson, M. S. Bernstein, E. Gerber, A. D. Shaw, J. Zimmerman, M. Lease, and J. Horton. The future of crowd work. In CSCW, 1301-1318, 2013.

10. Aniket Kittur, Boris Smus, Robert Kraut: CrowdForge: crowdsourcing complex work. CHI Extended Abstracts 2011: 1801-1806.

11. Anand Pramod Kulkarni, Matthew Can, Bjrn Hartmann: Turkomatic: automatic recursive task and workflow design for mechanical turk. CHI Extended Abstracts 2011: 2053-2058.

12. D. B. Martin, B. V. Hanrahan, J. O'Neill, and N. Gupta. Being a turker. CSCW, 224-235, 2014.

13. J. Rogstadius, V. Kostakos, A. Kittur, B. Smus, J. Laredo, and M. Vukovic. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. ICWSM 2011.

14. Elena Simperl, Barry Norton, Denny Vrandecic: Crowdsourcing Tasks within Linked Data Management. COLD 2011.