Making Sense of Data Workers' Sense Making Practices

Jiali Liu, Nadia Boukhelifa, James Eagan

Making Sense of Data Workers' Sense Making Practices

Jiali Liu
LTCI, Telecom ParisTech,
Université Paris-Saclay
46, rue Barrault
Paris, France
jiali.liu@telecom-paristech.fr

Nadia Boukhelifa
INRA, Université Paris-Saclay
1 av. Brétignières, 78850,
Thiverval-Grignon, France
nadia.boukhelifa@inra.fr

James R. Eagan
LTCI, Telecom ParisTech,
Université Paris-Saclay
46, rue Barrault
Paris, France
james.eagan@telecom-paristech.fr

Abstract
Data workers are non-professional data scientists who engage in data analysis activities as part of their daily work. In this position paper, we draw on our past experience in studying their data analysis processes and workflows, and the tools we built to support sensemaking. We describe our background as computer scientists and our multidisciplinary approach. Finally, we conclude with open questions and research directions, and argue for more research into the challenges faced by data workers.

Author Keywords
Data science, sensemaking, visualisation, visual analytics

Introduction
Data science activities are not only conducted by professional data scientists, but also widely performed by experts from various domains in both commercial and academic organizations [1, 4]. These domain experts may work under various job titles such as "researcher", "historian", and "medical surgeon". For example, an intelligence analyst at a police department may need to make sense of incident/crime reports everyday and discover patterns, trends, and any top issues in the city [1]. An HCI researcher may need to analyse participants' data collected in experiments to verify the impact of a novel technique.

For such domain experts, data analysis may not be their primary job, but making sense data is necessary to support decision-making and inform future directions. We call these non-professional data scientists, or domain experts, “data workers” [4].

In this position paper, we describe some of our past work on understanding the sensemaking practices of data workers in the face of uncertainty. We describe more generally our approach and background, and close with open research questions drawn from our experience.

Despite growing research attention to the sensemaking process [6, 5, 7], we have few insights into the data workers' work practices in real-world. For example, the tasks they engage in, the tools they use in the analysis, the strategies they deploy to cope with different problems, and the challenges they may face. These aspects all have an influence on their sensemaking activities.

Unlike well-trained data scientists, data workers have various level of skills, including in programming and the tools they use. They have varying technical expertise. They may use diverse tools from traditional paper-based settings (such as post-it notes), to general data management tools (e.g. Excel, MATLAB, computational notebooks, etc), to more domain specific tools (e.g. bioinformatics analysis tools, Jigsaw, GIS tools, etc).

Background & Approach

Information visualisation is a fundamentally interdisciplinary field, requiring a certain understanding of social and cognitive science, design, and computer science, as well as often an understanding of the data application domain. We are researchers in information visualisation and more generally human-computer interaction, with a background in computer science. One co-author is a researcher in an applied laboratory for agronomics research, while the other two are in an engineering school.

Our approach focuses on the intersection of sensemaking as a human cognitive process enabled through appropriate tools. As such, it is necessary and important to understand and join both the theory of how people make sense of data and the tools they use to perform it, both of which may have a strong interaction on the other as human and tool co-adapt [9]. We thus use a combination of various qualitative methods and tool-building to better understand human needs and the roles of various tools in satisfying them.

Most existing data analysis tools provide the building blocks to process, analyse, visualise, and present data across the various stages of the visualisation pipeline. They rarely, however, provide explicit support for the iterative refinement of the broader sensemaking process. It is up to the user to decide how to combine these tools, record previous state, and retrieve previous configurations.

Our goal is to provide better tools that integrate such support for the sensemaking process, from generally supporting the kinds of questions analysts might ask to providing explicit support for managing the different kinds and sources of, say, uncertainty that might arise.

Our work often builds on qualitative user studies that aim to understand how different kinds of data workers make sense of their data with a goal of actually building tools that address their results. In prior work, for example, we classified the questions students asked when evaluating different visualisation tools with an aim to understand what kinds of low-level tasks analysts perform [2]. More recently, we have focused more specifically on data workers and how they think about and manage the various kinds of uncertainty that arise in their work [4].

Our goal is to use these findings to inform the creation of tools that can specifically address identified practices, needs, and shortcomings of existing tools. For example, our work on low-level analytic tasks [2] has directly influenced the design of Jigsaw [11], among others.

In other work, we take a tool-first approach, where we build tools to help us better understand the needs of analysts within a specific context. For example, our work on evolutionary visual exploration aims to create better human-machine partnerships in exploratory visualisation of large data spaces [3].

Sensemaking Under Uncertainty

Sensemaking in the context of data analysis pertains to the iterative process of collecting, organising, exploring and reporting findings to answer specific research questions [10]. Extracting sense from data also involves cognitive tasks such as hypothesising, interpreting, and making inferences. We recently studied sensemaking under uncertainty as carried out by data workers, and identified common cognitive and data manipulation tasks. In terms of low level analytical tasks, data workers *acquire* and *manipulate* data, *characterise* various types and sources of uncertainty, *reason* about the data and the uncertainty, and *present* their findings [4]. The uncertainty characterisation process is key to sensemaking, and is carried out in various degrees of formality according to the data workers' work domains and skills. This can range from loose annotations, to statistical forms of uncertainty, to formal models. On a much higher level, data workers adopt a number of uncertainty coping strategies, often in combination, aiming to *understand* this uncertainty, to *minimise* it, to *exploit* (as an additional valuable source of information), or even to *ignore* it.

Our work thus far subscribes to a larger body of "data-centric" sensemaking research. More recently, however, there is growing interest on making sense of algorithms and models [12], complex and unfamiliar visualisations [8], as well as making sense of the sensemaking process itself, which is one of the objectives of this workshop.

Open questions & research directions

We envision opportunities for further research in sensemaking addressing the following open challenges:

Data workers have access to a wide range of tools from simple annotations to sophisticated computational packages. How do we study sensemaking when carried out in close partnership with machine learning?

Data workers deploy a variety of sensemaking strategies to cope with uncertainty. How do we create sensemaking tools that support different analysis strategies?

Data workers need to revisit their findings, and to share them with others. How do we facilitate collaborative sensemaking in a environment with differing skills and expertise?

The reasoning process in sensemaking encapsulates tasks that are a result of generation of thoughts, insights and decisions. These results are currently not easily exploitable. How do we record and make sense of these sensemaking processes?

As tool designers, how do we know whether we succeeded in supporting sensemaking? Can we come up with evaluation metrics that help us assess the efficacy of our sensemaking support tools? Should we evaluate based on time, error, insights, or do we need novel evaluation methodologies?

REFERENCES

1. Y. a. Kang and J. Stasko. 2012. Examining the Use of a Visual Analytics System for Sensemaking Tasks: Case Studies with Domain Experts. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec 2012), 2869–2878. DOI: http://dx.doi.org/10.1109/TVCG.2012.224

2. Robert Amar, James Eagan, and John Stasko. 2005. Low-Level Components of Analytic Activity in Information Visualization. In *InfoVis '05: Proceedings of the IEEE Symposium on Information Visualization*. IEEE Computer Society, Washington, DC, USA, 111–117. DOI: http://dx.doi.org/10.1109/INFOVIS.2005.24

3. N. Boukhelifa, W. Cancino, A. Bezerianos, and E. Lutton. 2013. Evolutionary Visual Exploration: Evaluation with Expert Users. In *Proceedings of the 15th Eurographics Conference on Visualization (EuroVis '13)*. The Eurographs Association & John Wiley & Sons, Ltd., Chichester, UK, 31–40. DOI: http://dx.doi.org/10.1111/cgf.12090

4. Nadia Boukhelifa, Marc-Emmanuel Perrin, Samuel Huron, and James Eagan. 2017. How Data Workers Cope with Uncertainty: A Task Characterisation Study. In *CHI '17: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3645–3656. DOI: http://dx.doi.org/10.1145/3025453.3025738

5. Garrett Grolemund and Hadley Wickham. 2014. A Cognitive Interpretation of Data Analysis. *International Journal of Statistics* 82, 2 (2014), 184–204.

6. Harlan Harris, Sean Murphy, and Marck Vaisman. 2013. *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*. O'Reilly Media, Inc.

7. E. Kandogan, A. Balakrishnan, E. M. Haber, and J. S. Pierce. 2014. From Data to Insight: Work Practices of Analysts in the Enterprise. *IEEE Computer Graphics and Applications* 34, 5 (Sept 2014), 42–50. DOI: http://dx.doi.org/10.1109/MCG.2014.62

8. S. Lee, S. H. Kim, Y. H. Hung, H. Lam, Y. A. Kang, and J. S. Yi. 2016. How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking. (Jan 2016). DOI: http://dx.doi.org/10.1109/TVCG.2015.2467195

9. Wendy E. Mackay. 1990. *Users and Customizable Software: A Co-Adaptive Phenomenon*. Ph.D. Dissertation. Massechusetts Institute of Technology.

10. Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. (2005), 2–4.

11. John Stasko, Carsten Görg, and Zhicheng Liu. 2008. Jigsaw: Supporting Investigative Analysis Through Interactive Visualization. *Information Visualization* 7, 2 (April 2008), 118–132. DOI: http://dx.doi.org/10.1145/1466620.1466622

12. Kanit Wongsuphasawat, Daniel Smilkov, James Wexler, Jimbo Wilson, Dandelion Mané, Doug Fritz, Dilip Krishnan, Fernanda B. Viégas, and Martin Wattenberg. 2018. Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow. *IEEE Trans. Vis. Comput. Graph.* 24, 1 (2018), 1–12.