Making Sense of Data Workers’ Sense Making Practices

Abstract
MAX 150 words

Author Keywords
Authors’ choice; of terms; separated; by semicolons; include commas, within terms only; required.

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCl)]: Miscellaneous; See [http://acm.org/about/class/1998/]: for full list of ACM classifiers. This section is required.

Introduction
...

Data Workers
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 needs to make sense of incident/crime reports everyday and discover patterns, trends, and any top issues in the city [1]. An HCI researcher needs to analysis the participants data collected in experiments to test the impact of a novel technique. A GIScience expert may...
need to analysis the crowd-sourced GPS logs that walkers publish online to plan for a new update to the company’s database[4].

For this group of domain experts, data analysis is not their primary job, but to make sense of the data that they gathered is a necessary approach to support decision-making and inform the following directions. These non-professional data scientists, or domain experts, are coined as “data workers” by Boukhelifa et al.[4].

Despite the growing research attention to data analysis sensemaking process[6] [5][7], we have few insights into the data workers’ work practices in real-world. For example, the tasks they engaged 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 models.

Unlike well-trained data scientists, data workers have various level of skills, including programming and tools using. They may have limited technical training background. The tools that they use vary from traditional paper-based settings(like post-it notes), to general data management tools(excel, matlab, computational notebooks, etc), to more domain specific tools(bioinformatics analysis tools, Jigsaw, GIS tools, etc).

The expertise of data workers, accrued over years of experience, plays an important role in the specific data analysis tasks[8]. Domain knowledge and professional judgments contribute throughout the sensemaking process, from acquiring datasets, drawing on intuitions, building mental models, to finding arguments, authoring and deciding. Different coping strategies have been deployed to handle different task problems. For example, in terms of managing analysis under uncertainty, Boukhelifa categorizes four types of methods based on interviews with 12 data workers: to understand, minimize, exploit or ignore uncertainties [4].

**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 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 [10]. 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, analyze, 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 (FIXME: USED SOME METHOD) aimed to understand what kinds of low-level tasks analysts perform when using visualisation tools [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 [12], 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**

Sensemaking in the context of data analysis pertains to the iterative process of collecting, organising, exploring and reporting findings to answer specific research questions [11]. Extracting sense out of data also involves cognitive tasks such as hypothesising, interpreting, and making inferences. We recently studied sensemaking under uncertainty as carried out by data workers. 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 sensemaking research that is focused on raw data as the main study subject. More recently, however, there is growing interest on making sense of algorithms and models [13], complex and unfamiliar visualizations [9], as well as making sense of the sensemaking process itself, which is one of the objectives of this workshop.

**Open questions/research directions**

We envisage opportunities for further research in sensemaking addressing the following open challenges: [what would be a good order for these?]

- 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 an 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 sense making support tools? Should we evaluate based on time, error, insights, or do we need novel evaluation methodologies?

REFERENCES


