Social Data Analysis in Co-located Environments

Petra Isenberg

University of Calgary

Department of Computer Science

2500 University Dr. NW

Calgary, AB T2N 1N4

petra.isenberg@ucalgary.ca

Sheelagh Carpendale

University of Calgary

Department of Computer Science

2500 University Dr. NW

Calgary, AB T2N 1N4

sheelagh@ucalgary.ca

Abstract

Our research focus is on supporting social data analysis in co-located settings. In this position paper, we outline our interests in the topic and highlight past research on social data analysis. We believe that the design of software for social data analysis in different temporal or spatial settings (co-located/distributed and synchronous/asynchronous) share some similar challenges and that research in each area can help to inform the other.

Keywords

Social Data Analysis, Collaboration, Co-located Work

Introduction

With the rapid growth in size and complexity of datasets, the practicality of an individual analyzing an entire dataset may become unrealistic. Instead, the expertise to analyze and make informed decisions about these information-rich datasets is often best accomplished in a collaborative setting. Collaborative data analysis combines the expertise and information processing power of many people.

The term "social data analysis" was first defined as a version of exploratory data analysis that relies on *social interaction* as source of inspiration and motivation [6]. It has been applied to scenarios in which a distributed

group of people join in asynchronous discussion and analysis of data primarily motivated by the social nature of the communication around the visualizations ([1],[4],[5],[6]). The goal can be to either reach a large audience by, for example, making information visualization tools publicly accessible and providing means for an open discussion about views, discoveries, and datasets (e.g. [1],[5],[6]) or to reach a specific audience with which to share and analyze the data (e.g. [4]).

In contrast to the work in distributed settings, our focus is on supporting small groups collaborating around information visualizations in a co-located setting. We believe that both distributed and co-located collaboration around data share some similar goals and challenges. In general, a collaborative information analysis setting can support social exchange about the data discoveries. Group members can discuss, debate, argue, and negotiate interpretations of the viewed data. The ability for this type of social exchange about visualizations of data is what unites both distributed and co-located data analysis scenarios under the term social data analysis. Our interest in attending this workshop is to gain a better understanding of how people work *together* over information displays. Ultimately, we want to further our understanding of the requirements for co-located information visualization design. We believe that we can learn a lot about collaboration around data by looking at different analysis scenarios and we hope that by sharing our insights into the challenges and requirements of co-located social data analysis, others will also benefit.

Our experiences with social data analysis

We will briefly talk about our experience with distributed social data analysis and then focus on our past and current research on co-located synchronous collaboration around data.

Distributed social data analysis

In [4], we introduced KeyStrokes, an artistic visualization of typed messages that captures and encodes aspects of an individual's unique typing style.

Figure 1: An image of two people's typed messages with the KeyStrokes system.

Figure 1 shows a comparison of two visualizations showing two people's unique typing of the same written message. We conducted an evaluation in form of a questionnaire to investigate people's motivation for later use of our tool and to find the appropriate platform to deploy KeyStrokes. Our study revealed that the

social interactions among viewers of the visualizations are motivations for people to use the tool, making it essentially a social data analysis tool. One of our participants specifically confirmed this finding by stating: "[The tool would be] a lot of fun to use, especially in a group setting." In contrast to systems like sense.us [1] or Many Eyes [5], we found that KeyStrokes would be used in a smaller, close community of people who already are familiar to each other. Many participants reported that they would use KeyStrokes for personalizing typed messages when corresponding to friends and family but would not use it for professional communication. Visualizations would be sent to friends or family who could then decode and analyze the visualizations to get some understanding about a person's typing. We plan to embed KeyStrokes in a specific communication environment (e.g. an email client) and conduct further studies on how the tool will be used and accepted in the group setting.

Co-located Collaboration

Recently, we shifted our research focus towards supporting small groups of co-located people in synchronous data analysis using large shared displays.

One of the challenges of research on co-located collaboration around data is that we do not have a clear understanding of how people collaboratively analyze information in a co-located setting and how information visualizations are used in this process. Knowledge about this process is important to enable effective social interaction around data that is presented in a digital form. To shed some light on this question, we conducted an exploratory study of the analysis processes of groups in contrast to individuals in a non-digital setting [3]. We wanted to observe people performing

analysis tasks without confounds of a digital system in order to observe how people would naturally interact with each other and share visualizations during data analysis. Our results identified a set of eight analysis processes that were common to all our participant groups. We found similarities between our processes and those described in previous work, yet, in contrast to some previous work no common temporal order of processes emerged. We hypothesize that this finding was possible because we based our observations in non-digital environments. Had we observed participants working with a digital collaborative system instead, we would have seen participants performing only those interactions provided by the digital system. Further evaluations will have to be conducted to see how the analysis process changes when interactive visualizations are used.

Figure 2 Two people using our system for social data analysis in a co-located collaborative setting.

Starting with the findings from our study, we wanted to expand our knowledge about the requirements for the design of co-located collaborative information visualization systems. In [2] we present a set of challenges and requirements compiled from a review of literature in the Computer Supported Cooperative Work (CSCW) field, information visualization design guidelines, and results from studies about collaborative visualization settings. In the same article we also presented a new system for small group work around hierarchical data comparison based on these guidelines. An example of our system in use can be seen in Figure 2. Our system includes support for multi-user input, flexible layout of information in the workspace to aid mental model formation, flexible representation changes, and support for multiple collaboration styles. Yet, this implementation and the assembly of design guidelines is only a beginning and the set of guidelines will hopefully expand as our research on collaboration around information visualization continues. Throughout this work we noted that there are several shared challenges for the design of co-located or distributed visualization systems, for example the difficulties of view vs. value changes in data representations, the integration of findings from multiple representations of a shared dataset, or the creation of histories of an exploration process. Research on social data analysis, whether different in temporal or spatial interaction of participants, can further our understanding of how interfaces, visualizations, and interaction techniques should be designed to address the data analysis needs of groups.

Acknowledgements

References

[1] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. In *Proc. of CHI*, pages 1029–1038, New York, NY, USA, 2007. ACM Press.

[2] Petra Isenberg, Sheelagh Carpendale. Interactive Tree Comparison for Co-located Collaborative Information Visualization. IEEE Transactions on Visualization and Computer Graphics (Proceedings Vis/ Infovis), 12(5), Sept./Oct. 2007. To appear.

[3] Petra Neumann, Anthony Tang, Sheelagh Carpendale. A Framework for Visual Information Analysis. Technical Report 2007-87123. University of Calgary. 2007.

[4] Petra Neumann, Annie Tat, Torre Zuk, and Sheelagh Carpendale. KeyStrokes: Personalizing Typed Text with Visualization. In *Proc. EuroVis*, pages 43–50, Aire-la-Ville, 2007. Eurographics.

[5] Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. Many Eyes: A Site for Visualization at Internet Scale. *IEEE TVCG (Proc. Vis / Infovis 2007)*, 12(5), September/October2007.

[6] Martin Wattenberg. Baby Names, Visualization, and Social Data Analysis. In John Stasko and Matt Ward, editors, *Proc. InfoVis*, pages 1–7, Los Alamitos, CA, USA, 2005. IEEE Computer Society.