Supporting the Sensemaking Process: An Analytic Provenance Approach ?or? Provenance Data Visualization for Support Sensemaking

Phong H. Nguyen

Associate Professor Kai Xu

Professor William Wong

School of Science and Technology

Middlesex University

This dissertation is submitted for the degree of

Doctor of Philosophy

May 2016

Contents

Chapter 1
Introduction

Introduce:

– sensemaking and explain why supporting it is important and challenging

– vis and its capability of supporting sensemaking

1.1 Research Problem and Approach

Overall, this research aims to *examine how to support users in their sensemaking process.* Pirolli and Card [7] describe sensemaking as a process consisting of two loops: the *foraging loop*, which involves searching, extracting and organizing information; and the *sensemaking loop*, which involves building schemas, generating and testing hypotheses, and presenting the outcome. Schematization plays an important role in this model: connecting the foraging loop and the sensemaking loop. It is a crucial step in converting raw evidence to rational explanations. Pirolli and Card suggest that the schematization process should be supported by a computer-based tool that organizes raw evidence into small-scale stories about typical topics or in answer to typical questions (e.g., who, what, when, where, why, how). This suggestion is aligned with a recent empirical study by Kang, Görg and John Stasko [5]. It shows that all of the participants who performed the sensemaking task well spent considerable time and effort in *organizing their collected information.* Their organizational schemes were flexible: a *timeline* of related events, a *map* connecting locations that a person has been to, and a *diagram* showing relationships among suspicious targets.

One approach to provide such support is through **analytic provenance** – an area of research focusing on understanding a user's reasoning process through the study of their interactions with a visualization tool [6]. More specifically, during the process in which a user solves a sensemaking task using a visualization tool, their interactions and discoveries – we refer to both of them as *provenance data* – are

captured and visualized to provide support back to the sensemaking process itself as summarized in Figure 1.1.

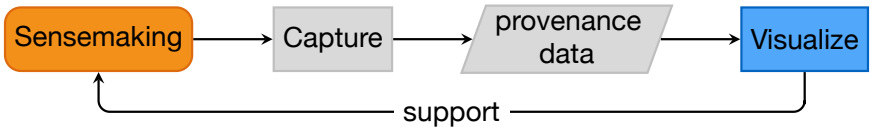

Figure 1.1: A pipeline of supporting sensemaking through analytic provenance.

In this thesis, we focus on the **visualization** part of this process:

How to design interactive visualizations of provenance data to support sensemaking?

The research is driven by enabling users to discover the answers to the aforementioned questions (who, what, when, where, why, how) of the captured provenance data.

1. **when?** this is to help users to analyze the **temporal** relationship of their provenance data. Current timeline visualization has these limitations
   - no or very simple layout which is cluttered and space-inefficient
   - designed for presenting a known story rather than interactively constructing a hidden one
2. **who/what?** grouping information related to a person or a topic together allows users to have a better understanding of their provenance data. It is more useful if such **thematic** information can be visualized together with the temporal information. Currently, this is still an open challenge. These three principles of Gestalt's laws of grouping are typically employed to visualize set relations on a timeline:
   - similarity: use colors or shapes to indicate sets – not powerful
   - proximity: not space-efficient
   - uniform connectedness: cluttered
3. **why?** discover and visualize **semantic/causal** relationships of user's provenance data is a step towards answering this question. After a long period of working on the sensemaking task, the user may get lost. At a low level, they may fail to find the information they discovered before. At a higher level, they may not know where they are in the context of the overall task, and may not know where to continue. Existing visualizations do not fully support users to overcome this *disorientation* problem.

We choose not to address the **where** question because spatial information is not always available and relevant to the task. If locations are named, we can consider them as *themes* and use the methods for **who/what** questions to analyze them. We do not address the **how** question; however, combining the answers to **when** and **who/what** questions may help improve understanding about it.

We take a user-centered design approach in developing the solutions to the three aforementioned questions. For each question, we elicit the design requirements either by conducting a user study or drawing from the literature. Visual encoding and interaction are designed to meet those requirements and are implemented into a working prototype. Finally, an empirical study is conducted to validate if the tool supports users as intended.

The analytic provenance approach that we take is general; however, we need to choose specific domains/tasks to demonstrate its application. Sensemaking does not only happen when using a visualization system to solve a "serious" intelligence analysis task. Actually, it often happens in our life such as when using a web browser to find information to select the most suitable smartwatch to buy. Therefore, to demonstrate the wide application of our visualization techniques, we target both intelligence analysis tasks using a visualization system and everyday sensemaking tasks using a web browser.

1.2 Thesis Organization and Contributions

Chapter 2 provides an overview of sensemaking, visualization and analytic provenance.

Chapter 3 reviews related work of visualizing provenance data to support sensemaking.

Chapter 4 presents a compact yet aesthetically pleasing timeline visualization technique that enables users to interactively construct a temporal schema from provenance data.

Chapter 5 extends Chapter 4 to present a timeline visualization technique that can effectively show both temporal and set information of provenance data. This technique can also be used for more general data.

Chapter 6 discusses the requirements, design, implementation and evaluation of a visualization system that enables users to have an overview of their sensemaking process, to organize their provenance data in such a way that aids their understanding about the task, and to communicate their findings at different levels of granularity.

Chapter 7 describes a general approach that combines analytic provenance and visualization to shorten the transcription and coding steps in qualitative data analysis, and a tool to demonstrate its application.

Finally, Chapter 8 concludes the thesis with a discussion on its contributions and future research directions triggered from this work.

Chapter 2
Background

2.1 Sensemaking

- Overview of sensemaking
    - Pirolli and Card's model
    - Data–frame model

2.2 Visualization and Visual Analytics

- Overview of vis, scivis, infovis, visual analytics
    - Show how they're used to support sensemaking

2.3 Analytic Provenance

- overview of provenance, other types of provenance, then analytic provenance
    - modeling of analytic provenance: 4-level model by Gotz and Zhou and other vis task/action taxonomies
    - pipeline: capture - visualize - reuse and a summary of capture and reuse. [Visualize stage is the main focus on this thesis thus should have more discussion than other parts. Move it to a separate chapter to make sections balanced and also avoid one extra nested level of numbering.]

Chapter 3

Review of Provenance Data Visualization for Supporting Sensemaking

3.1 Temporal Visualization

Overview of different techniques to visualize temporal data

- Focusing on timeline
- Timeline for sensemaking

3.2 Timeline Visualization with Set Relations

Set visualization

- Timelines with set relations

3.3 Causal/Relational Visualization

VA systems providing 'reasoning workspace'

Chapter 4
Timeline Visualization of Provenance Data

Timeline visualization is an important tool for sensemaking. It allows analysts to examine information in chronological order and to identify temporal patterns and relationships. However, many existing timeline visualization methods are not designed for the dynamic and iterative nature of the sensemaking process and the various analysis activities it involves. In this chapter, we introduce a novel timeline visualization, SchemaLine, to address these deficiencies.

4.1 Introduction

The importance of timeline visualization in supporting sensemaking.

The current limitation

- similarity: use colors or shapes to indicate sets – not powerful
- proximity: not space-efficient
- uniform connectedness: cluttered

The contribution

- a novel design for an interactive timeline that groups notes into schema determined by the analyst,
- an algorithm to automatically generate a compact and aesthetically pleasing visualization of these schema on the timeline, and
- a set of fluid interactions with the timeline to support the sensemaking activities defined in the Data-Frame model.

4.2 Requirements

[not in the paper yet! List requirements – drawn from the literature – that SchemaLine need to support]

4.3 Design

Visual representation of individual events and schemas

4.4 Algorithm

Describe the algorithm that produces the compact layout of schemas and events.

4.5 Sensemaking with SchemaLine

Discuss how SchemaLine's interactions support sensemaking activities in the Data–frame model. [the requirement for this support should be mentioned in the Requirements section.]

4.6 Application

Discuss the integration of SchemaLine into INVISQUE. SchemaLine receives data input as user notes of INVISQUE and the linking between schemas and index-cards.

4.7 Evaluation

A case study with 3 participants (different backgrounds) to use INVISQUE+SchemaLine to solve an intelligence analysis task using VAST Challenge 2011 dataset. Report how they used the tool, how the tool might help them.

4.8 Conclusion

Chapter 5
Timeline Visualization with Set Relations of Provenance Data

The timeline visualization in Chapter 4 cannot show events belonging to multiple sets. In this chapter, we introduce a novel timeline visualization technique, TimeSets, that helps make sense of complex temporal datasets by showing the set relationships among individual events. TimeSets visually groups events that share a topic, such as a place or a person, while preserving their temporal order. It dynamically adjusts the level of detail for each event to suit the amount of information and display estate.

5.1 Introduction

The current limitation

- no or very simple layout which is cluttered and space-inefficient
- designed for presenting a known story rather than interactively constructing a hidden one

The contribution: TimeSets

- clearly shows the events within a set over time and their relationships with other sets;
- dynamically adjusts the level of details of each event to suit the amount of information and display estate;
- uses color gradient backgrounds for events belonging to multiple sets and curved set outlines to emphasize its grouping.

5.2 Design

Visual representation of individual events and sets

5.3 Layout

Describe the layout algorithm.

5.4 Evaluation

A controlled experiment with 30 participants to compare TimeSets and the state-of-the-art set visualization technique – KelpFusion.

5.5 Case Study 1: Publication Data

Shows an application of TimeSets to publication data of 200 articles and 8 sets.

5.6 Case Study 2: VAST Challenge 2014

Shows an application of TimeSets to tweets data.

Also use TimeSets to show user findings [not done yet! currently, findings are shown as a node-link diagram.]

5.7 Conclusion

Chapter 6
Understanding Relationship of Provenance Data

Very often, users get lost when solving a complicated task using a big dataset over a long period of exploration and analysis. They may forget what they have done, are not aware of where they are in the context of the overall task, and do not know where to continue. In this chapter, we introduce a tool, *SenseMap*, to address these issues in the context of *browser-based online sensemaking*.

6.1 Introduction

Limitation: Existing approach – graphical browser history – only provides a static overview of the browsing process

Contribution:

1. A user study exploring how users search, manage and synthesize online information for their daily work activities; and a series of workshops followed up to generate design questions and formulate solutions.
2. A visual analytics tool SenseMap supporting browser-based online sensemaking addressing all the elicited requirements.
3. A user evaluation exploring how SenseMap is used in a naturalistic work setting and a process model derived from the data analysis.

6.2 Design Research and Requirements

Describe the study to elicit requirements

6.3 SenseMap

Describe the tool addressing all the requirements

6.4 Evaluation

We conducted a user-centered evaluation of SenseMap in order to:

- Evaluate its effectiveness in providing the desired support for searching, managing and synthesizing information through our collect, curate, communicate process model;
- Identify significant features relating to user behaviors, interactions and outcomes in the use of this new tool; and
- Describe any processes that these behaviors, interactions and outcomes might infer.

6.5 Conclusion

Chapter 7
Case Study: Understanding the Sensemaking Process

In this chapter, we consider a specific sensemaking task: *a qualitative study to understand user's sensemaking process*. This is an essential task because understanding the sensemaking process allows building effective visual analytics tools to make sense of large and complex datasets. Currently, it is often a manual and time-consuming undertaking to comprehend this: researchers collect observation data, transcribe screen capture videos and think-aloud recordings, identify recurring patterns, and eventually abstract the sensemaking process into a general model. We propose a general approach to facilitate such a qualitative analysis process, and introduce a prototype, *SensePath*, to demonstrate the application of this approach with a focus on *browser-based online sensemaking*.

7.1 Introduction

Current limitation: transcription and coding are very time-consuming.

Contribution

1. A general approach combining the strength of analytic provenance and visual analytics to understand user's sensemaking process. This approach can be potentially applied to other qualitative research in HCI beyond sensemaking.
2. A qualitative study and a participatory design session to understand characteristics of qualitative research on sensemaking.
3. A visual analytics tool SensePath to demonstrate the general approach. It supports the transcription and coding of the observation data of online sensemaking tasks.

4. A qualitative user evaluation that demonstrated the effectiveness of the general approach and the tool SensePath.

7.2 Design Research and Requirements

Discuss how we approach the problem and elicit requirements to support.

7.3 SensePath

Describe the tool interface

7.4 Evaluation

We conducted two user-centered evaluations: the first one is to understand how SensePath is used by an experienced qualitative researcher and to identify opportunities for improvement, and the second one is to discover whether SensePath has any advantages compared to a traditional method.

7.5 Conclusion

References

[1] Bostock, M., Ogievetsky, V., and Heer, J. (2011). D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309.

[2] Dou, W., Jeong, D. H., Stukes, F., Ribarsky, W., Lipford, H. R., and Chang, R. (2009). Recovering Reasoning Processes from User Interactions. *IEEE Computer Graphics and Applications*, 29(3):52–61.

[3] Gotz, D. and Zhou, M. X. (2009). Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55.

[4] Jankun-Kelly, T. J., Ma, K.-L., and Gertz, M. (2007). A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):357–369.

[5] Kang, Y.-A., Görg, C., and Stasko, J. (2011). How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 17(5):570–583.

[6] North, C., Chang, R., Endert, A., Dou, W., May, R., Pike, B., and Fink, G. (2011). Analytic provenance: process+interaction+insight. In *ACM Transactions on Computer-Human Interaction*, pages 33–36. ACM.

[7] Pirolli, P. and Card, S. (2005). The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. In *Conference on Intelligence Analysis*.

[8] Russell, D., Jeffries, R., and Irani, L. (2008). Sensemaking for the rest of us. *CHI workshop on Sensemaking*.