MORGAN & CLAYPOOL PUBLISHERS

Semantic Interaction for Visual Analytics

Inferring Analytical Reasoning for Model Steering

Alex Endert

SYNTHESIS LECTURES ON VISUALIZATION

Niklas Elmqvist & David Ebert, *Series Editors*

Synthesis Lectures on Visualization

Editors
Niklas Elmqvist, *University of Maryland*
David S. Ebert, *Purdue University*

Synthesis Lectures on Visualization publishes 50- to 100-page publications on topics pertaining to scientific visualization, information visualization, and visual analytics. Potential topics include, but are not limited to: scientific, information, and medical visualization; visual analytics, applications of visualization and analysis; mathematical foundations of visualization and analytics; interaction, cognition, and perception related to visualization and analytics; data integration, analysis, and visualization; new applications of visualization and analysis; knowledge discovery management and representation; systems, and evaluation; distributed and collaborative visualization and analysis.

Semantic Interaction for Visual Analytics: Inferring Analytical Reasoning for Model Steering
Alex Endert
2016

Design of Visualizations for Human-Information Interaction: A Pattern-Based Framework
Kamran Sedig and Paul Parsons
2016

Image-Based Visualization: Interactive Multidimensional Data Exploration
Christophe Hurter
2015

Interaction for Visualization
Christian Tominski
2015

Data Representations, Transformations, and Statistics for Visual Reasoning
Ross Maciejewski
2011

A Guide to Visual Multi-Level Interface Design From Synthesis of Empirical Study Evidence
Heidi Lam and Tamara Munzner
2010

Semantic Interaction for Visual Analytics: Inferring Analytical Reasoning for Model Steering

Alex Endert

www.morganclaypool.com

ISBN: 9781627054713 paperback
ISBN: 9781627052917 ebook

DOI 10.2200/S00730ED1V01Y201608VIS007

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON VISUALIZATION

Lecture #7
Series Editors: Niklas Elmqvist, *University of Maryland*
David S. Ebert, *Purdue University*
Series ISSN
Print 2159-516X Electronic 2159-5178

Semantic Interaction for Visual Analytics

Inferring Analytical Reasoning for Model Steering

Semantic Interaction for Visual Analytics

Inferring Analytical Reasoning for Model Steering

Alex Endert
Georgia Tech

SYNTHESIS LECTURES ON VISUALIZATION #7

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

User interaction in visual analytic systems is critical to enabling visual data exploration. Interaction transforms people from mere viewers to active participants in the process of analyzing and understanding data. This discourse between people and data enables people to understand aspects of their data, such as structure, patterns, trends, outliers, and other properties that ultimately result in insight. Through interacting with visualizations, users engage in sensemaking, a process of developing and understanding relationships within datasets through foraging and synthesis.

This book discusses a user interaction methodology for visual analytic applications that more closely couples the visual reasoning processes of people with the computation. The methodology, called ***semantic interaction***, affords user interaction on visual data representations that are native to the domain of the data. These interactions are the basis for refining and updating mathematical models that approximate the tasks, intents, and domain expertise of the user. In turn, this process allows model steering without requiring expertise in the models themselves—instead leveraging the domain expertise of the user. Semantic interaction performs incremental model learning to enable synergy between the user's insights and the mathematical model.

The contributions of this work are organized by providing a description of the principles of semantic interaction, providing design guidelines for the integration of semantic interaction into visual analytics, examples of existing technologies that leverage semantic interaction, and a discussion of how to evaluate these techniques. Semantic interaction has the potential to increase the effectiveness of visual analytic technologies, and opens possibilities for a fundamentally new design space for user interaction in visual analytic systems.

KEYWORDS

user interaction, visual analytics, model steering, visualization

Acknowledgments

I appreciate the feedback and comments on drafts of this manuscript from Ross Maciejewski, Remco Chang, and Niklas Elmqvist. I would like to thank Doug Bowman, Scotland Leman, Richard May, and Francis Quek for serving on my Ph.D. committee and providing valuable feedback that helped create much of the work presented in this manuscript. Finally, I'd like to thank Chris North for his continued guidance, advice, and encouragement that made this research possible.

Alex Endert
September 2016

Contents

CHAPTER 1

Introduction

Visual analytics helps people make sense of data by combining information visualization with data analytics through interactive user interfaces. The current data-driven era emphasizes the importance of usable, yet powerful tools that help people understand their data, and thus the world around us. Visual analytic applications serve an important role in these situations, as they create usable interfaces for people to discover insights into their data.

Performing visual data analysis involves a blend of data analytics and human reasoning. The appropriate blend of these two parts needs to be explicitly balanced through design and evaluation. To achieve this balance, people and computers co-reason about data using each of their processing abilities, with the goal of giving people insights. There has been much emphasis on the computational and analytic components of visual analytic systems. The same is true of the visual side; novel visualizations continue to be created. As a result, researchers and developers have created advanced analytics that are faster and more accurate, and able to uncover more complex forms of phenomena and patterns from data. Novel visualization techniques have also been researched and developed that allow people to perceive and understand new and intricate data characteristics, as well as results of analytic models. User interaction, on the other hand, has seen fewer innovation and design alternatives.

The integration of analytic models into visual analytics continues to be an open research challenge. While the visual analytics community advocates for approaches to data analysis that include human perception and cognition, there are advantages to incorporating analytic models into the process. Some data analysis tasks are more well-suited for automation, while other tasks that are perhaps less structured and defined are better for humans to perform. For example, computing clusters from data is more efficiently performed by computation, as long as the parameters of the clustering algorithm are known by the user. In contrast, generating questions, hypotheses, or stories about potentially interesting insights may be better suited for human reasoning. The understanding of which tasks during analysis are better suited for computation or cognition is an open area of research, yet an area well-suited for visual analytics.

User interaction is a growing focus area for visual analytics. The understanding of how to integrate people into the computational processes and visualization methods used for visual analytics is increasing in importance. Interaction transforms people from mere viewers of visualizations into active participants in the visualization and analysis processes. People can contribute to this process in different and important ways. Their expertise and knowledge about a domain can allow them to focus the computation toward more relevant or important features of the data.

Alternatively, in order to understand the data being shown, exploring different aspects and views of that data might be helpful.

Semantic interaction is an approach to user interaction that couples exploratory interactions with updating and steering computational models. The premise of semantic interaction is to create user interaction techniques that more closely couple this cognitive processing and reasoning of humans with the computational processes and models used in visual analytics. The goal is to create systems that optimize the balance between human and machine effort for data analysis. For such systems, user interaction is the means through which this coupling takes place.

1.1 THE ROLE OF VISUAL ANALYTICS IN A DATA-DRIVEN ERA

Visual analytics is a science based on supporting sensemaking of large, complex datasets through interactive visual data exploration [69]. The success of such systems hinges on their ability to combine capabilities of statistical models, visualization, and human intuition—with the goal of supporting the user's analytic process. Through interacting with the system, users are able to explore possible connections, investigate hypotheses, and ultimately gain insight. This complex and personal process is often referred to as sensemaking [57].

In the digital era, an increasing number of physical phenomena are being digitized and quantified through sensors and data transformations. This data is stored in ever-growing databases, against which scientists, practitioners, and others pose questions about the world and the phenomena in it. This can reveal new insights into existing areas of science, and also help people working in other domains such as security, finance, business, and others. For many domains and tasks, the mere presence of data presents opportunities to answer existing scientific questions using new, data-driven methods.

Visual analytics occupies an important role in the myriad of data science tools that help people make sense of data. For example, visual analytics can help people understand and contextualize the results of automated computation. Often, simply finding the answer using one or more automated methods leaves people questioning the accuracy of that answer, or how much they should be able to trust it [99]. In these situations, visual analytics can help people explore the processes and decisions of how these automated processes came up with the results. Visualization is a powerful mechanism for explaining the results of artificial intelligence, machine learning, predictive modeling, and other automation. It can help people understand why models came up with the answer they did, why other alternatives were not returned, what the confidence of these results are, etc.

Alternatively, visual analytics can also serve as the user interface to help people explore and analyze data more freely. People can ask questions, test hypotheses or prior expectations about the data, and engage in other, more exploratory tasks. For the latter, the visual analytic techniques provide people with the interfaces needed to tune and adjust the analytic models that compute on the data, and are often used to generate the visualizations contained in them. Finally, visual analytics can help overcome some of the potential drawbacks in automated computation (e.g., helping machine learning models delineate signal from noise), as well as human-in-the-loop techniques (e.g., helping people formulate the questions to ask of their data) [98]. The growth and evolution of visual analytics plays a pivotal role in the advancement of data science.

1.2 SEMANTIC INTERACTION

Semantic interaction [25] is an approach to user interaction that enables coupling between computational models and human reasoning. Semantic interactions are designed to be performed on visual objects that are well-understood by analysts in the domain of the data. Sequences of semantic interactions are systematically interpreted to produce updated and steered mathematical models, which in turn drive updated visualizations and foster analytic discourse. Semantic interaction interprets sequences of user interactions as data structures that capture semantically meaningful information about the user so that systems can use the information during analysis, and also analyzing a user's process after analysis.

Visual analytic techniques that make use of semantic interaction let people interact with the visual representations themselves to impart their domain expertise onto the computational models. For example, people may have domain expertise that suggests a strong similarity between two or more data items. This can be conveyed by visually grouping those items. In response, the system learns the similarity function that can describe this relationship, and applies the learned model to the rest of the data to produce a new visualization. This iterative process continues as people explore the data and gain new insights.

To date, semantic interaction has focused on enabling direct manipulation of spatializations, which are two-dimensional views of high-dimensional data such that similarity between information is represented by relative distances between data points (e.g., a cluster represents a collection of similar information) [66]. Other applications of semantic interaction include additional visual metaphors, different computational models, and other domains. Some of these are discussed in the later sections.

The interactions people perform are small, piecewise realizations of the cognitive analysis process of analysts. Thus, *semantic interaction* is a systematic process by which the analytical reasoning or meaning (i.e., the *semantics*) of the interactions are interpreted by the application. As a result, the cognitive processes of reasoning about data are more closely connected with the computational processes of machines.

To illustrate the concept of semantic interaction, an existing prototype ForceSPIRE (shown in Figure 1.1) is used throughout the following chapters. ForceSPIRE is a visual analytic prototype incorporating semantic interaction for analysis of text document collections represented in a spatialization [25]. Semantic interactions in ForceSPIRE include repositioning documents, highlighting text, searching, and annotating documents. When users perform semantic interactions in the course of their reasoning process, the system incrementally updates a keyword weighting scheme in accordance with the user's analytical reasoning (Table 4.1). The learned weighting scheme emphasizes relevant keyword entities within the dataset and adjusts the layout of the spatialization accordingly. Thus, the goal of ForceSPIRE is to automatically steer the spatialization based on the user's interaction with a subset of the information. It is used to test and evaluate the balance between people and machines in visualizing text documents spatially, via multiple analytic models.

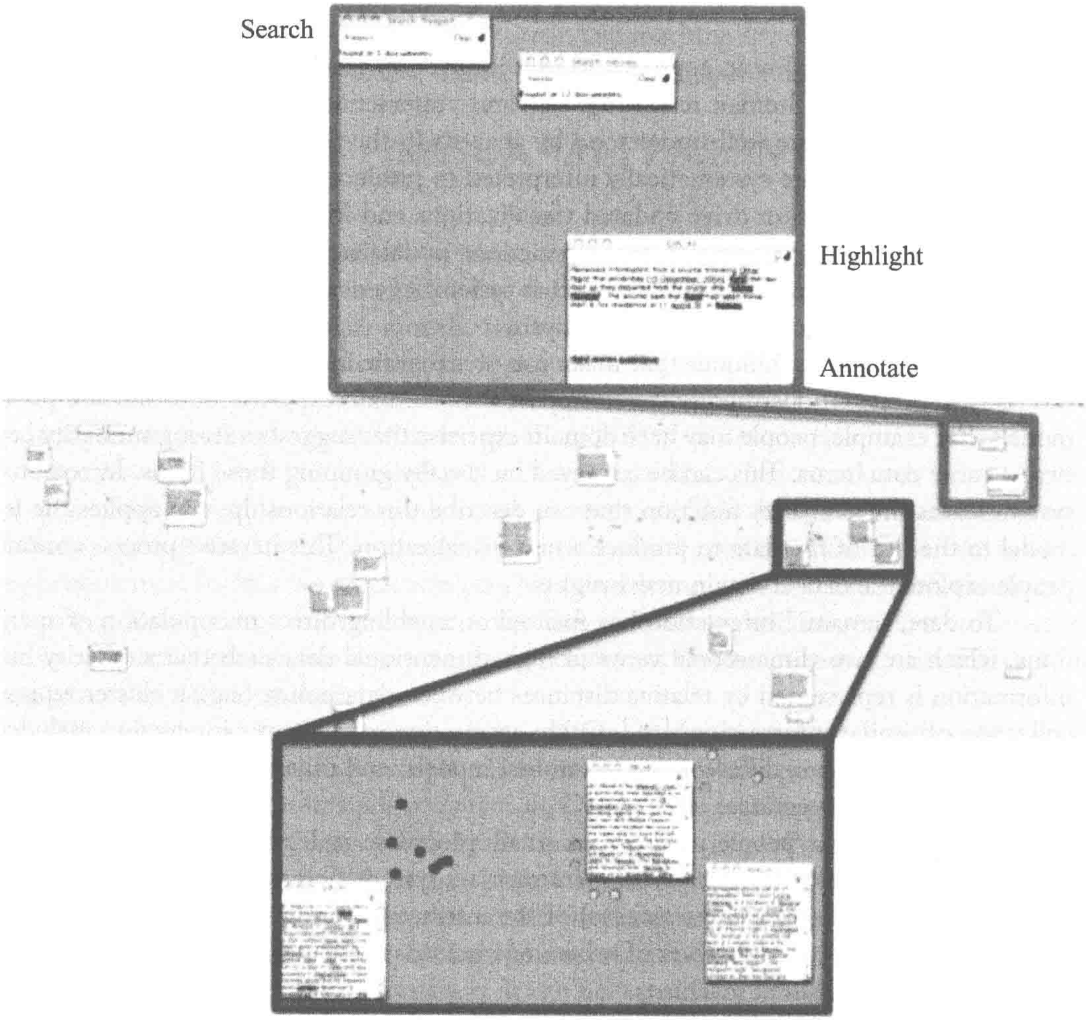

Figure 1.1: A scaled-down screenshot of ForceSPIRE taken on the large, high-resolution display used in this study (two zoomed views shown). Users can search, highlight, annotate, and reposition documents spatially. Documents can be shown as minimized rectangles, as well as full detail windows.

1.3 OUTLINE

Chapter 2 starts with an overview of the fundamental concepts that ground semantic interaction for visual analytics. It consists of topics including model steering, mixed-initiative systems, analytical reasoning, sensemaking, and creating models from captured user interactions from visualizations.

Chapter 3 discusses the value of the spatial visual metaphor, specifically for sensemaking. It covers some of the basic studies and research performed which grounds much of the spatial adjustment and manipulation of data underpinning semantic interaction.

Chapter 4 covers the design guidelines for semantic interaction. These are described in context of applying this form of user interaction to visual analytic systems with a variety of dataset and analytic models. Many of these design guidelines are exemplified in applications shown in subsequent chapters. This section also describes a general framework for semantic interaction.

Chapter 5 provides a select set of visual analytic applications that utilize different forms of semantic interaction. The coupling between the user interactions and the model steering techniques are discussed for each.

Chapter 6 presents current approaches for evaluating systems that utilize semantic interaction. The implicit model steering that is at the core of semantic interaction requires novel methods for evaluation. In addition, the captured temporal models approximating user interest opens opportunities for novel user evaluation techniques for visual analytics.

Chapter 7 discusses semantic interaction more broadly in the context of visual analytics. Additionally, it mentions some of the research directions and challenges raised. Chapter 8 provides an overall summary and conclusion.