

CHI 2011 Proposal for Workshop on Analytic Provenance

**POC: Wenwen Dou, UNC Charlotte
wdou1@uncc.edu**

Proposal

Proposal: Analytic Provenance: Process + Interaction + Insight

Chris North

Virginia Tech
2202 Kraft Dr.
Blacksburg, VA 24060
north@cs.vt.edu

Remco Chang

Tufts University
161 College Ave.
Medford, MA 02155
remco@cs.tufts.edu

Alex Endert

Virginia Tech
2202 Kraft Dr.
Blacksburg, VA 24060
aendert@cs.vt.edu

Wenwen Dou

UNC Charlotte
9201 University City Blvd,
Charlotte, NC 28223
wdou1@uncc.edu

Richard May

Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Richard.may@pnl.gov

Bill Pike

Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Bill.pike@pnl.gov

Glenn A. Fink

Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Glenn.fink@pnl.gov

Abstract

Visual analytics research reveals a new opportunity at the intersection between three areas of research: the Analytic Process, Interaction, and Insight. Studying the analytic process, we have learned about sensemaking, analytic methods, procedures, and analytic tasks. User interactions with analytic tools have investigated the mechanics and exploratory operations of the user. Studying the generation of insight through the user interactions, we have gained knowledge on the cognitive aspects of analysts' mental models. The question now is: How does *interaction* in the *analytic process* produce *insight*? It appears the missing link of analytic provenance lies in the intersection of these three areas, rather than in any single one.

Timing

We believe CHI 2011 is an ideal time and venue to hold a workshop on Analytic Provenance. There is a growing recognition of the fundamental importance of interaction in the analytic process, causing a shift in thinking among researchers in the field about the definition of analytic provenance. Thus, we are at a critical stage at which it would be very beneficial for researchers to come together and discuss the points of view and consolidate the future research agenda. This

Copyright is held by the author/owner(s).

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

ACM 978-1-4503-0268-5/11/05.

would encourage focused effort and enable a flourishing of research on this important topic.

Impact

We suggest that the topic of analytic provenance can be roughly broken down into five areas based on existing literature: perceive, capture, encode, recover, and reuse. However, part of the goal of this workshop is for the attendees to discuss their viewpoints and to determine a more comprehensive research agenda. The existing published work on analytic provenance can be organized into the following topics:

Perceive: We contend that the process of analytic provenance begins with understanding what the user sees. This is critical for making sense of the user's subsequent interactions.

Capture: Capturing of interactions is the heart of most work relating to analytic provenance. Various systems have been studied that demonstrate the ability to capture and log user interactions within a tool.

Encode: Representing the captured interactions in a generalizable way is challenging. Systems exist which can encode interactions, but most are limited to the tool in which the user operates.

Recover: Once the user's interactions have been captured and encoded (stored), the challenge becomes making sense of the interactions. Solutions can come in either an automated or manual (human) form.

Reuse: One of the final goals of analytic provenance is to be able to automatically reapply the relevant user knowledge through interaction capturing, encoding, and

recovery to a new data set of problem domain. This challenge remains largely unsolved, as much of the focus in the community is currently on the previous topics.

Our workshop will explore these key topics, as well as discuss any relevant views expressed in the submitted position papers.

Tangible Outcomes

- Produce a research agenda publication
- Collaborative research proposals
- Host a special issue journal with paper submissions
- Initiate recurring workshop on topic
- Host an on-line resource to promote continued collaboration

Workshop Organization

Pre-Workshop

The organizing committee will create an Analytic Provenance community on the VACommunity.org website detailing the focus area of the workshop, the schedule, and submission instructions. Participants will be asked to submit a short, 2-page purpose statement outlining how their research relates to the focus of this workshop. Upon finalizing the participants, the organizing committee will initiate and maintain contact with each participant via email and the website. This website will remain online after the workshop to foster continued collaboration and research in addition to documenting this first critical workshop.

The intended size of this workshop is 20-25 participants. The selection procedure for participants will be based upon the acceptance of a peer reviewed

2-page purpose statement that each participant (or group of participants) will be expected to submit (organizing committee, invited participants, and advisory board exempt).

Workshop Schedule

This is a 2-day workshop, formatted as follows:

Day 1, Morning

Introductions: Each participant will briefly introduce themselves, focusing on a single key topic/issue/challenge relevant to the workshop. From these, a select few will drive the breakout sessions.

Analysts' Processes and Needs: Invited analysts from a variety of domains will describe their work, processes, and challenges. Participants will take note of key topics, and have a chance to ask questions.

Afternoon

Current Systems/Approaches Analysis and Evaluation: From the position papers, we will select key systems or approaches that will be presented/demonstrated. Participants will be shown the functionality of each tool/technique to be used in the following breakout session.

Day 2, Morning

During this session, participants will break into small groups, each working with one of the tools/techniques demonstrated. Each group will solve a sample dataset, focusing on how the functionality of each tool/technique affects the analytic process. This session will conclude with an open-floor discussion with short breaks in between.

Afternoon

Identify Future Research Agenda: Focused discussion and small breakout groups aimed at establishing a research agenda for analytic provenance. This discussion will frame the research agenda publication (one of the outcomes of the workshop). After a short break, we will have the results of the session and concluding remarks.

Evening

Dinner/Social Event: We will provide dinner for participants to socialize and network.

After the Workshop

At the end of the workshop, we will produce a publishable research agenda with clear goals on how to improve in each focused area. In addition, to sustain this research effort, we will also discuss the potential of establishing recurring workshops or symposiums and special issues of journals where research in this area can be published. Finally, with the help of the advisory board, we will identify funding opportunities and broader impact of this research and seek support from government funding agencies.

Workshop Organizers

Chris North is an Associate Professor of Computer Science at Virginia Tech, where he is head of the Laboratory for Information Visualization and Evaluation. His research is in the areas of human-computer interaction, information visualization, large high-resolution displays, and visualization evaluation methods. His current work examines how analytic insight relates to analytic process and user interaction. He is particularly interested in how large high-

resolution displays can be used to better support capturing, viewing, and reusing analytic provenance.

Remco Chang is an assistant professor at Tufts University whose research includes information visualization, visual analytics, and computer graphics. His ongoing collaboration with Bank of America on risk analysis gave him exposure to how financial analysts perform fraud detection, and led him to study methods for capturing and reusing these analytical procedures.

Alex Endert is a Ph.D. student at Virginia Tech. His work focuses on visual analytics and visualization, and how these apply to large displays. Ongoing collaboration with intelligence analysts at Pacific Northwest National Laboratory has led him to pursue how large displays enable fundamentally different interactions, namely spatial interaction, organization, and querying.

Wenwen Dou is a Ph.D. student at University of North Carolina at Charlotte. Her research is in the areas of visual analytics and human-computer interaction. Her current work focuses on exploring the relationship between analytic provenance and user interaction. She is also working on developing visualization systems and interaction techniques for program officers at NSF to make sense of research trends and science policies.

Richard May is a chief scientist at the Pacific Northwest National Laboratory and Director of the National Visualization and Analytics Center (NVAC). His research focus for the past several years has been in visual analytics and interaction methodologies. His particular interest is the logical and physical aspects of interacting with information for analytical tasks using

visual analytic techniques. He manages both research and development projects as well as outreach programs to government, industry, and academia.

William Pike is a senior research scientist at Pacific Northwest National Laboratory where he focuses on visual analytics, analytic methods, and visualization for data-intensive computing. He has developed techniques for capturing analysis processes and recording insight in visualization software, and he works extensively with end users to integrate support for analytic workflows into visualization software.

Glenn A. Fink has been a Senior Research Scientist at Pacific Northwest National Laboratory in Richland, Washington since 2006. Dr. Fink specializes in computer security, visualization, and human-centric computing, centering computer systems' design and function around the needs and abilities of people. Dr. Fink is conducting ongoing research at PNNL in adaptive computer security systems with a human-centric point of view.

Extended Abstract

Analytic Provenance: Process + Interaction + Insight

Chris North

Virginia Tech
2202 Kraft Dr.
Blacksburg, VA 24060
north@cs.vt.edu

Remco Chang

Tufts University
161 College Ave.
Medford, MA 02155
remco@cs.tufts.edu

Alex Endert

Virginia Tech
2202 Kraft Dr.
Blacksburg, VA 24060
aendert@cs.vt.edu

Wenwen Dou

UNC Charlotte
9201 University City Blvd,
Charlotte, NC 28223
wdou1@uncc.edu

Richard May

Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Richard.may@pnl.gov

Bill Pike

Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Bill.pike@pnl.gov

Glenn A. Fink

Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Glenn.fink@pnl.gov

Abstract

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. One key aspect that separates visual analytics from other related fields (InfoVis, SciVis, HCI) is the focus on analytical reasoning. While the final products generated at from an analytical process are of great value, research has shown that the processes of the analysis themselves are just as important if not more so. These processes not only contain information on individual insights discovered, but also how the users arrive at these insights. This area of research that focuses on understanding a user's reasoning process through the study of their interactions with a visualization is called *Analytic Provenance*, and has demonstrated great potential in becoming a foundation of the science of visual analytics. The goal of this workshop is to provide a forum for researchers and practitioners from academia, national labs, and industry to share methods for capturing, storing, and reusing user interactions and insights. We aim to develop a research agenda for how to better study analytic provenance and utilize the results in assisting users in solving real world problems.

Keywords

Analytic provenance, user interaction, visual analytics, visualization

Copyright is held by the author/owner(s).

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

ACM 978-1-4503-0268-5/11/05.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI).

Introduction

Understanding a user’s analytic reasoning process when using a visual analytics system has become an important research topic in the visual analytics community. Central to the mission of the visual analytics research agenda [1], this research aims at understanding how a user interacts with a visual interface to perform analytical tasks. With such an understanding, researchers and developers can design better interfaces that assist reasoning flow, enable knowledge sharing, and eventually support human-computer mixed initiative systems [1].

Although recent research has shown that a user’s reasoning process can be retrieved through examination of a user’s interaction history [2], there is little agreement on how to best capture a user’s interactions, store the user history, or retrieve the user’s reasoning process. Researchers in various domains have designed and implemented proprietary mechanisms that are suitable for their domains (such as automatic tutorial generation [3], scientific visualization [4], network detection [5], etc.), but it is largely unclear how the success of one system can be applied to a different system in an unrelated domain. The goal of this workshop is to bring these researchers and practitioners together to share their experiences, and discuss what steps are necessary for developing a deeper understanding of analytic provenance as both a theory and a practice.

Background

A central precept of visual analytics is that the development of human insight is aided by interaction with a visual interface, and the steps that a user takes to discover insights are often as important as the final product itself [6]. The key to the research of analytic provenance is the belief that by capturing a user’s interactions with a visual interface, some aspects of the user’s reasoning processes can be retrieved. In practice, we propose that the research of analytic provenance can be examined in five interrelated stages: *perceive*, *capture*, *encode*, *recover*, and *reuse*.

Perceive

In order to correlate a user’s interactions with a visualization to her reasoning process, the research must begin with understanding how the data is presented to the user. As shown by Dou et al., combining the visual representation with the interaction history can disambiguate “why” a user performs certain interactions [2]. Since the user’s interaction can only begin after perceiving the visualization of data, the analytic provenance research also needs to start with the understanding of how information is perceived by the user.

Capture

As the user interacts with visualization, the series of interactions can be considered as a linear sequence of actions. The most common application of this concept is the use of “undo” and “redo” buttons that are available to most computer software today [7]. However, such information is often insufficient in representing the user’s reasoning process. Researchers have shown that additional semantic information is necessary to adequately represent a user’s analysis

process [6]. Such semantic information can be directly annotated by the user [8], modeled based on task analysis [9], or correlated with the visualization elements [2], but identifying the most appropriate representation remains an open challenge [10].

Encode

Encoding refers to the process of describing the captured provenance in predefined formats. While many systems implicitly have their own encoding schema for capturing analytic provenance for specific tasks and domains, few generalizable schemas exist. Researchers have attempted using XML [11], declarative pattern language [5], logic-programming [12], and dynamic scripts [13], but in most cases these schemas only record the “how”, but not always the “why”. By using these schemas, the user can reapply interaction, but the semantic meanings behind these steps are often unclear.

Recover

Once the user’s provenance has been captured and encoded, the challenge becomes making sense of the provenance. As noted by Jankun-Kelly et al., history alone is not sufficient for analyzing the analytical process with visualization tools [11]. Often, there are relationships between the results and other elements of the analysis process which are vital to understanding. While some of the relationships have been shown to be recoverable through manual inspection [2], whether the same can be done using automated techniques is still an open question.

Reuse

One important goal of the research in analytic provenance is to be able to automatically reapply a

user’s insights to a new data or domain. As noted earlier, most systems that are successful at encoding a user’s interactions have mechanisms that allow for the reapplication of the interactions within the same system [5, 11, 12, 13]. However, in most analytical environments, analysts often utilize multiple tools simultaneously which renders the use of existing methods inadequate. A more comprehensive and cohesive encoding, recovering, and reusing process is therefore necessary to support the analysts in their natural working environments.

Key Questions to Discuss

Although various user interaction logging technologies exist, we still lack a fundamental understanding of how user interactions can be captured and transformed to insights, and how a visual analytics system can utilize such insights to assist a user in performing future analytical tasks. A number of issues remain open for investigation, and this workshop aims to bring together researchers to examine these issues critically based on their experiences in studying user interactions and provenance capturing. Using the five stages of analytic provenance, these questions can be categorized into:

- Perceive: How is information visually presented to the user that affects the user’s reasoning process?
- Capture: What types of user interactions should be captured, and how much semantic information should be included based on a user’s task?
- Encode: How should the system store the recorded user interactions? Can the encoded interaction be shared across multiple systems?

- Recover: Based on captured interactions, how can a user's reasoning process be recovered? Can the recovery be done automatically (by a computer)?
- Reuse: How can a visual analytics system apply what it has learned about a user's reasoning process to assist the user in performing future analyses? Can the learned reasoning process be applied to other tasks and other systems?

Expected Participation

We have received a significant amount of interest from diverse groups of researchers in academia, government labs, and industry who have been investigating the relationships between process, interaction, and insight. We therefore expect these participants to bring their expertise in computer graphics, scientific visualization, information visualization, visual analytics, sensemaking, decision making, and HCI to this workshop. With such different backgrounds and interests, we believe that a significant and impactful research agenda can be developed that will be used as a roadmap of future research in the theory and practice of analytic provenance.

References

- [1] Thomas, J. and Cook, K. Illuminating the path. National Visualization and Analytics Center, 2004.
- [2] Dou, W., Jeong, D.H., Stukes, F., Ribarsky, W., Lipford, H.R., and Chang, R. Recovering reasoning process from user interactions. *IEEE Computer Graphics and Applications* (2009). 29(3): 52-61.
- [3] Grabler, F., Agrawala, M., Li, W., Dontcheva, M., Igarashi, T. Generating Photo Manipulation Tutorials by Demonstration. *SIGGRAPH* (2009), 66:1-66:9.
- [4] Bavoil, L., Callahan, S., Crossno, P., Freire, J., Scheidegger, C., Silva, C., and Vo, H. Vistrails: enabling interactive multiple-view visualizations. *Proc. IEEE Visualization* (2005), 135-142.
- [5] Xiao, L., Gerth, J., and Hanrahan, P. Enhancing visual analysis of network traffic using a knowledge representation. *Proc. IEEE Visual Analytics Science and Technology* (2006), 107-114
- [6] Pike, W.A., Stasko, J., Chang, R., and O'Connell, T.A. The science of interaction. *Information Visualization* (2009), 8(1): 263-274
- [7] Heer, J., Mackinlay, J., Stolte, C., and Agrawala, M. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics* (2008), 14(6):1189-1196.
- [8] Shrinivasan, Y. and van Wijk, J. Supporting the analytical reasoning process in information visualization. *Proc. ACM CHI* (2008), 1237-1246.
- [9] Fink, G.A., North, C.L., Endert, A., and Rose, S. Visualizing cyber security: Usable workspaces. *Proc. Visualization for Cyber Security* (2009), 45-56.
- [10] Gotz, D. and Zhou, M. Characterizing users' visual analytic activity for insight provenance. *Proc. Visual Analytics Science and Technology* (2008), 123-130.
- [11] Jankun-Kelly, T., Ma, K. and Gertz, M. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics* (2007), 13(2): 357-369.
- [12] Garg, S., Nam, J., Ramakrishnan, I., and Mueller, K. Model-driven visual analytics. *Proc. IEEE Visual Analytics Science and Technology* (2008), pp. 19-26.
- [13] Kadivar, N., Chen, V., Dunsmuir, D., Lee, E., Qian, C., Dill, J., Shaw, C., Woodbury, R. Capturing and supporting the analysis process. *Proc. IEEE Visual Analytics Science and Technology* (2009), 131-138.

Call for Participation

Workshop on Analytic Provenance: Process + Interaction + Insight

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. One key aspect that separates visual analytics from other related fields (InfoVis, SciVis, HCI) is the focus on analytical reasoning. Research has shown that the processes of the analysis themselves are just as important as the resulting products. These processes contain information on insights and how the users arrive at these insights. This area of research that focuses on understanding a user's reasoning process through the study of their interactions with visualization is called *Analytic Provenance*, and is a potential foundation of the science of visual analytics. The goal of this workshop is to provide a forum for researchers and practitioners to share methods for capturing, storing, and reusing user interactions and insights. We aim to develop a research agenda for how to better study analytic provenance and utilize the results in assisting users in solving real world problems.

The workshop (see <http://www.vacommunity.org/AnalyticProvenanceWorkshop>) will involve short presentations and demonstrations by participants, feedback from practicing analysts, observations of usage of various analytic tools that illustrate provenance, and breakout sessions for discussing research agendas.

Submissions to the workshop should be sent to Alex Endert (aendert@cs.vt.edu) in the form of a position statement (maximum 2 pages). Submissions should describe work that is ongoing, with either demonstrations of working systems or evaluations of the role of analytic provenance in visual analytics problems.

If accepted, at least one author of the paper will have to register for the conference and the workshop.