Uncertainty Propagation and Trust Building in Visual Analytics

Dominik Sacha, Hansi Senaratne , Bum Chul Kwon, *Member, IEEE* and Daniel A. Keim, *Member, IEEE*

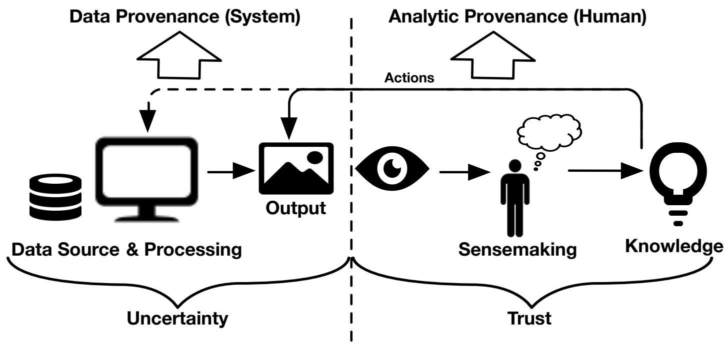

Figure 1: The role of uncertainty and trust along the visual analytics process related to data and analytic provenance. Uncertainty builds up from data source to the system output that is perceived by users. Human user's sensemaking involves trust in order to arrive at valid knowledge in the end. Such activities in visual analytics systems can be captured as data and analytic provenance.

Abstract

Visual analytics combines human and machine abilities to generate new knowledge from data. Within this process, *uncertainty* often plays an important role in hindering the sensemaking process and analysis tasks. On the machine side, uncertainty builds up from the data source level to the visual output. On the human side, these uncertainties often result in "lack of knowledge or trust" or "overtrust." Such human's biased interpretation can be resolved if we can measure uncertainties and users' trust at each stage and provide proper mitigation in time. We believe that we can achieve this by tracing data provenance and analytic provenance accurately and reflecting them on the system output. Therefore, our first goal is to identify the roles of uncertainty and trust along the entire visual analytics knowledge generation process. In addition, we aim to capture how uncertainty and trust can be derived from data and analytic provenance. In this workshop, we introduce a framework that describes the roles of uncertainty and trust, and introduce open research questions with potential solutions.

Keywords: Visual Analytics, Provenance, Sensemaking, Trust, Uncertainty, Knowledge Generation

1 Introduction

In the visual analytics process, users arrive at new knowledge after performing innumerable sensemaking activities. The goal of visual analytics is to foster effective collaboration between human and machine that improves the knowledge generation process. To succeed in this process, end users should be able to trust their knowledge generated from visual analytics. Analysts often can be blinded from uncertainties that are hidden in the black box of visual analytics systems, such as their data sources, preprocessing, analysis processes or visualizations. Therefore, it is crucial for users to be able to grasp the accurate estimation of uncertainties from visual analytics systems so that they can increase trust in their generated knowledge. The benefit of such demystified systems is to prevent analysts from over- or under-trusting the results of data analysis. On the human side, uncertainties often cause analysts to have low confidence in their findings. With such low confidence, users often fail to narrow down their answers for a hypothesis and can be confused with conflicting interpretations. On the other hand, analysts can sometimes overtrust outcomes without verification and understanding inherent uncertainties. Such overtrusted findings can easily lead to unwarranted and inaccurate knowledge at the end. Therefore, our goal is to externalize uncertainties as well as human users' confidence and trust at various states of analysis path. To achieve this, data provenance and analytic provenance should be captured and presented accurately. Many prior studies tackled parts of these problems. There exist provenance and knowledge management tools as well as visual analytics systems that present inherited uncertainties in system outputs. However, none of these systems can yet solve the entire problem from the machine side as well as the human side.

Recently, the field of analytic provenance emerged as an important discipline in visual analytics to analyze the human reasoning and sensemaking process [4]. There are several approaches that capture user's behavior during the analysis process (e.g., Dou et al., [2]; Gotz et al., [3]). For example, HARVEST [3, 6] tracks and analyzes human behavior in order to recommend visualizations and the visual layout of items that are under investigation. In addition, the knowledge generation process has been studied, especially with regards to the definition of knowledge and its pre-stages (e.g., Chang et al., [1]; Sacha et al., [5]). Wang et al., [7] presents a provenance-aware visual analytics system that tracks the data lineage which is also identified as a standardized geo-spatial data quality component among others. Some other approaches capture and feed these stages of knowledge (externalization) back into the system for further analysis (Wang et al., [8]).

2 Knowledge Generation Process

We take the knowledge generation model for visual analytics as a basis of our study [5]. The model defines and relates machine and human concepts for which we will consider uncertainty and trust aspects in Section 3. The machine side consists of the *Data*, *Visualization* and *Model* nodes as well as the mappings in between these nodes: *Visual Mapping*, *Model Building* and *Model-Vis Coupling*. The human part of the model consists of cognitive, reasoning and sensemaking concepts embedded in a three-loop framework. The *Exploration Loop* consists of *Actions* that a user performs with the systems' components and *Findings* which are defined as useful patterns, visual artifacts and model outputs that can be used for further analysis. In this loop, analysts perform low-level actions and observe findings until they start gaining meaningful interpretation from their findings. The humans then merge into the *Verification Loop* where analysts can test their *Hypotheses* with *Insights*. Hypotheses have to be validated through many verification and exploration cycles. The exploration loop is steered by the verification loop as new hypotheses are generated and tested by evidence from the exploration loop. The confirmed hypotheses become mature as they are refined and revised throughout the knowledge generation loop. This loop then creates the knowledge, which is defined as "justified belief".

3 Uncertainty and Trust in Visual Analytics

Uncertainty and trust play important roles in generating valid knowledge. In this section, we illustrate how uncertainty and trust interplay at each stage of the knowledge generation process. This illustration hints how data and analytic provenance can be captured and used to estimate uncertainty and trust.

Visual analytics systems use various techniques at individual stages to lower such uncertainties, but they often create new ones by doing so. At the beginning, the data source already has various inherent uncertainties stemming from the process of data collection and production. These types of errors transfer incomplete, logically inconsistent, and uncertain data to the next preprocessing stage. At the preprocessing stage, the system provide structure into data for analysis by using various methods, such as sampling, quantization, and interpolation. However, each of such computations sometimes add new variations in the data. Models may also introduce uncertainty when results are based on predictions, simulations or classifications. Visualizations sometimes illustrate these uncertainties with some techniques (e.g. transparency). However, such additional encodings may also introduce uncertainties if the visualization technique is not chosen appropriately.

In our definition, the role of "trust" is to facilitate or deter tranformation of findings with inherited uncertainty to the next level outcome, such as insights and knowledge. In other words, trust can be self-measured confidence in results. The knowledge generation process continues as follows. At every exploration or verification cycle, analysts stay in the loop until he believes that he collected enough confidence (refined). If the analysts can interpret findings and trust these insights with high confidence, then he arrives at new knowledge. A "lack of knowledge/trust" forces the human to increase his confidence though repeating the loop. This lack of knowledge refers to multiple possible interpretations for a finding (insights), to vague hypotheses and even to the users actions (e.g., does the user know what he is doing?) or findings (e.g., is the finding easy to spot?). If analysts "undertrust" findings, they may be stuck in one analysis loop, which could waste their time and energy. On the other hand, if analysts "overtrust" findings in analysis loops, then they could arrive at inaccurate conclusions. Figure 1 illustrates the relations and transformations from uncertainty to trust. It highlights that accurate estimation of uncertainties would increase trust in analysis results, which will result in fluid and effective visual analytic processes.

4 Research Questions and Future Work

In this section, we formulate novel research questions that can be investigated in the future.

First, we see the necessity to increase the awareness of uncertainties, often hidden in a black box for the analysts. Furthermore, they should be able to drill down into uncertainties that cascade all over the pipeline of visual analytics. As a solution, each component of visual analytics systems could automatically detect and present the amount of uncertainties with a global uncertainty measure that can clearly be communicated to the human users. This requires us to define and relate uncertainty measures for all techniques of data production, preprocessing, analysis models and visualizations. Such uncertainty measurements can be derived from data provenance.

Second, we should detect the level of confidence and the lack of trusted knowledge. To achieve this, capturing analytic provenance is very important. Based on the humans analytic paths, a system could detect the level of trust human users currently has for analysis outcomes. With the trust measurement, a system could provide a new analytic path that may ensure verification of insights for certain hypotheses. Such capabilities require accurate externalization of human knowledge and expertise at each step of the process. Besides a users actions, users' annotations could consist of direct feedback that indicate uncertainty (e.g., "this could be a possible interpretation for a spotted outlier but I'm not really sure"). The system should learn how to interpret such signals either from analytic provenance or direct feedback so that it can mitigate any potential issues. They key question is how to capture users' interaction with system artifacts and to predict users' states accurately.

Finally, the system should involve human users, but how can we involve the user in measuring these trust and uncertainties? A possibility could be to ask an analyst to provide direct inputs on confidence ratings for system outputs. Another idea is to nudge users to externalize their hypotheses, thoughts, findings, and insights. To do so, provenance components should offer capabilities to annotate, connect, and relate analysis results. Then, such externalized pieces of inforamtion can be analyzed for trust measurements. We envision a system that is able to detect the amount of uncertainty and trust at each analysis stage. Each conceptual state that appears in [5] on the human side should be part of a provenance component and enriched with uncertainty and trust measurements. In addition, such uncertainty and trust should be used to guide human users toward the valid knowledge generation.

The efficient integration and connection of these provenance, knowledge management, and visual analytic systems coupled with human interaction will improve the area of visual analytics to the next level. Our initial framework will provide a nice platform to gather ideas about how to achieve the goal. In our next step, we plan to derive uncertainty and trust measurements via computation of provenance data.

References

[1] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009.

[2] W. Dou, W. Ribarsky, and R. Chang. Capturing reasoning process through user interaction. *Proc. IEEE EuroVAST*, 2, 2010.

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

[4] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic provenance: Process+interaction+insight. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 33–36, New York, NY, USA, 2011. ACM.

[5] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visual Analytics Science and Technology 2014)*, 20(12), Dec. 2014.

[6] Y. B. Shrinivasan, D. Gotz, and J. Lu. Connecting the dots in visual analysis. In *IEEE VAST*, pages 123–130. IEEE, 2009.

[7] S. Wang, A. Padmanabhan, J. D. Myers, W. Tang, and Y. Liu. Towards provenance-aware geographic information systems. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, page 70. ACM, 2008.

[8] X. Wang, D. H. Jeong, W. Dou, S.-W. Lee, W. Ribarsky, and R. Chang. Defining and applying knowledge conversion processes to a visual analytics system. *Computers & Graphics*, 33(5):616–623, 2009.