EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2018)
K. Lawonn, N. Smit, L. Linsen, and R. Kosara (Editors)

Uncertainty of Visualizations for SenseMaking in Criminal Intelligence Analysis

M. J. Islam, K. Xu and B. L. W. Wong

Department of Computer Science, Middlesex University, London, UK

Abstract

Uncertainty in visualization is an inevitable issue for sensemaking in criminal intelligence. Accuracy and precision of adopted visualization techniques have got greater role in trustworthiness with the system while finding out insights from crime related dataset. In this paper, we have presented a case study to introduce concepts of uncertainty and provenance and their relevance to crime analysis. Our findings show how uncertainties of visualization pipeline influence cognitive biases, human awareness and trust-building during crime analysis and how provenance can enhance analysis processes that include uncertainties.

Categories and Subject Descriptors (according to ACM CCS): Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques - Interaction Techniques.

1. Introduction

Uncertainty is a necessary aspect of data analysis. During analysis numerous techniques are followed for allowing analysts to make observations and research claims with varying levels of authority. Failing to acknowledge uncertainties around such analysis task, dataset and analysis technique may lead to a cavalier and superficial data analysis: making faulty claims with confidence that may lead to poor decision making.

Crime analysis encompasses a range of data analysis activities. Many tasks, however, require analysts to study large collections of crime reports in order to identify aberrant or exceptional patterns of activity, identify new and emerging crime series, or sometimes suggest crime suspects that may be linked to a crime phenomenon. There are seldom concrete, single or certain approaches or techniques that can be taken at each of these stages and often solutions are found through serendipity instead of rules. These incorporate uncertainty into visual representation of data that may lead to erroneous insights due to inaccuracy that occurs through the pipeline of data processing. Brodlie et. al. [BOL12] has denoted this uncertainty as *Uncertainty of Visualization*. Such uncertainty becomes more problematic in crime solving domain as it may have negative consequences for innocent individuals. Current state of the art *Uncertainty of Visualization* differs with the concept *Visualization of Uncertainty* - which considers how we depict uncertainty specified with the data [BOL12] and on which lot of research works have been carried out to find techniques and develop tools. Current state of the art demands more work to find out causes and effects of uncertainty of visualization in criminal intelligence analysis.

2. Related Work

Many works have been carried out to visualize uncertainties through the various components of a system. Brodlie et. al [BOL12] has proposed a typology of visualization methods to handle uncertainties. To unpack the uncertainties those propagate through visual analytics system and their consequences on human perceptual and cognitive biases, Sacha et. al. [SSK*16] have proposed a knowledge generation model for visual analytics. To support uncertainty aware decision making in criminal intelligence Stoffel et. al. [SSEK15] describes a visionary system named as VAPD for Comparative Case Analysis (CCA). Decision making under uncertainty can lead to cognitive biases. Geoffrey et. al. [GD15] presents examples of situations where cognitive biases in visualization can occur during decision making process.

3. Crime Analysis Under Uncertainty

Visual analytic techniques raise challenges on trustworthiness of outcome while finding out the answers of five vital investigative questions [LPM01] - 5WH (Who, When, Where, What, Why, How) in criminal intelligence by using dynamically changing, incomplete, inconsistent dataset. If analysts are unaware of inherent uncertainties, they may waste their time by following wrong and uncertain leads. As found from the literature we have categorized such uncertainties that an analyst needs to deal with during processing and analyzing data according to following:

- **Personal Uncertainty**- processing obliges analyst to test the assumptions and hypotheses they have hitherto been operating with. The analyst has to ensure that the way in which information is organized enables a sober and unbiased evaluation of its

DOI: 10.2312/eurorv3.20181145

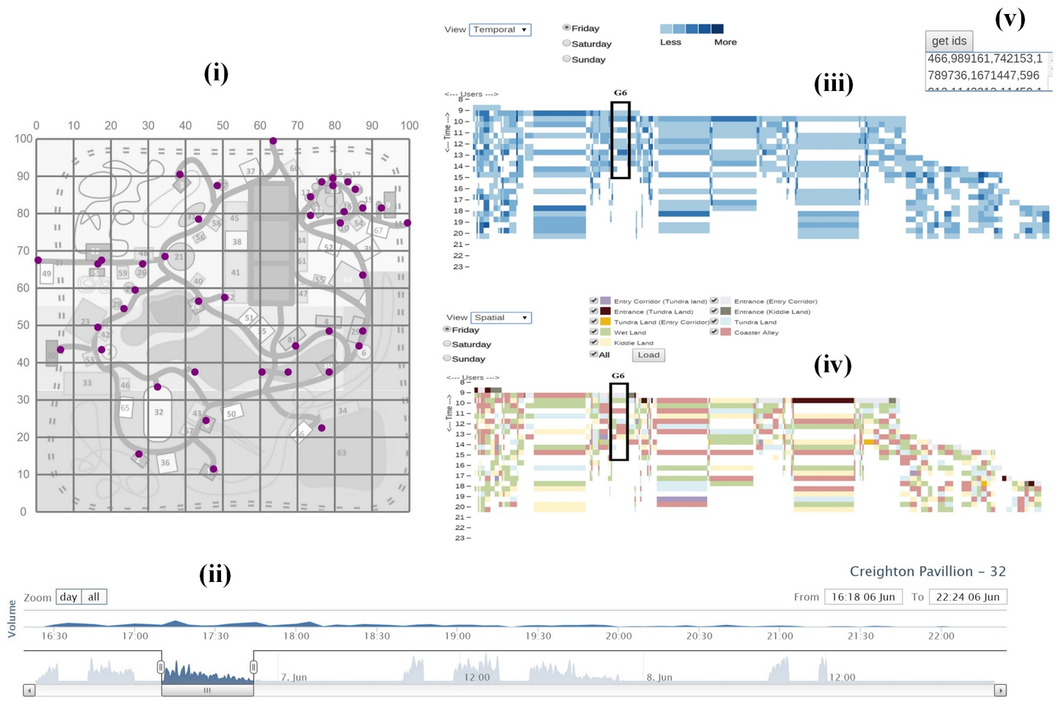

Figure 1: *Park visitor's check-in visualizations.* ***(i)*** *Purple colored dots are different check-in points,* ***(ii)*** *High-Chart view of overall check-ins temporally,* ***(iii)*** *Frequencies of temporal check-ins visualizations with color shades [less->more],* ***(iv)*** *Check-ins using spatial colors of different park areas as shown on park-map,* ***(v)*** *Filtered view of group check-in ids.*

contents. Errors introduced here can seriously affect any subsequent analysis.

- **Outcome Uncertainty**- the outputs of processing are inputs to analysis. Consequently, processing should be oriented to helping the analytic process. Similarly, if the steps to be taken during analysis are not clear, processing will be muddled.
- **Issue Uncertainty**- as with other steps in the cycle, understanding the issue at hand enables processing by providing the analyst with one or more concept models that can be applied to the structuring of information. Such models can be tacit or explicit in nature, technology driven or merely pen-and-paper representations to help the analyst filter and organize the data collected.
- **Decision Uncertainty**- processing encompasses the broadest range of possible activities. Given the resource constraints, analysts are often required to weigh the options available and determine which are most likely to generate new insights or ideas.
- **Information Uncertainty**- processing offers another opportunity to critically assess the information collected in terms of its reliability, accuracy and relevance, as well as to verify the quality of sources from which the data originated.
- **Visualization Biases**- analysts see patterns into data plots (e.g. on a scatter plot) when the data is in fact a random distribution. Two things are occurring here, i) the user is unaware that a random sample does not generate an even distribution of points on a simple scatter plot or in coin tossing, a fairly balanced sequence of heads and tails; and ii) humans are predisposed to finding patterns, even very insignificant ones such as three points in a row amongst hundreds of scattered points. This cognitive bias is one that has already been identified, and is in fact a visualization bias [GD15] rather than analytic.
- **Trust Building**- obviously, the chance of human error is highest when uncertainty is present in the system and the analyst is not aware of it, or mistakenly believes that there are no uncertainties. Uncertainty in visual analytics originates and propagates from the system that is the datasets, data model and visualizations and is then passed to the analyst as findings and insights are discovered, resulting in knowledge generation. Uncertainties affect human trust building processes using the knowledge generation model [SSK*16] for visual analytics.

Techniques that provide accurate estimates of uncertainties are therefore vital. By understanding the uncertainties, analysts better trust their acquired knowledge and can report findings with greater rigour and authority.

4. Case Study on a Criminal Situation

We conducted a case study on VAST Challenge 2015 [WCG*15] dataset to demonstrate how uncertainties may occur due to lack of appropriate technique of visualization. We considered problems of Mini-Challenge 1 and used it's available park visitor's 14.5M movements and 4.1M communications datasets for this case study. As part of initial processing we filtered out park visitor's check-ins dataset and visualized to have an understanding of the situation.

The Mini-Challenge 1 describes an incident of vandalism at Dino World (an amusement park) during a weekend (Friday, Saturday, Sunday) of June 2014. Park officials and law enforcement figures are interested in understanding just what happened during that weekend to better prepare themselves for future events. They are also interested in understanding how people move and communicate in the park, as well as how patterns changes and evolve over time, and what can be understood about motivations for changing patterns.

4.1. Visualization Paradigm

We developed two kinds of visualizations to show park visitor's check-ins over time i.e, `Temporal` and `Spatial`. Temporal visualization uses blue color shades to represent check-in frequencies of park visitors over time. Spatial visualization uses color codes of park map to visualize check-ins of park visitors at different areas over time.

All user check-ins at different areas have been visualized temporally into a user vs time matrix represented as $U \times T$ where `U=user` and `T=time` as shown in Figure 1.

4.2. Data Visualization and Exploration

Data visualization gives an idea of data structure and particularities. For example, it uncovers presence of any cluster into the data, whether the variables are correlated with each other, similarities among them or are there any outliers. From Temporal and Spatial views above few groups such as group `G6` have been identified based on their criteria of being the same group. The criteria includes same kinds of activities (i.e, check-in frequencies, movement patterns, check-outs etc.) through-out the whole day. These kinds of smaller/bigger group activities can be found quite a lot of time through-out the whole visualization.

4.3. Findings on Uncertainties

"Uncertainty is the dissimilarity between a given representation of reality and the known or unknown reality, where the unknown reality simply means you do not know what the reality actually is that you are representing" - the definition proposed by Plewe [Ple02] has similarity of fact into current visualizations. Due to incomplete representation of data there might be flaws in logic, vague or misapplied similarities to unrelated events resulting failures of imaginations to find a viable solution. We call this as `Determinacy Problem` which has got two types:

- **Spatial Determinacy** - exact location of the event happening.
- **Temporal Determinacy** - actual time of the event happening.

These determinacy problems lead to the uncertainty of space and time which means *"Don't know about when and where"*. As described into Mini Challenge-1 of VAST Challenge 2015 - *A news article was published in the newspaper on June 10, 2014 with the title "Mayhem at DinoFun World" - by Mako Harrison, staff reporter by saying that "The crime forced partial closure of DinoFun World and local police were on the scene shortly after the vandalism was discovered by park visitors. Security guards are being questioned to eliminate the possibility of an inside job. Creighton Pavilion-32 was closed and locked up tight before each show as stated by park Chief of Security Barney Wojciehowicz"*. This information gives a start of analyzing the crowd for initial understanding of the fact. Our spatio-temporal visualizations into Figure 1 show groups of people who checked-in together and got split after a while. Our visualizations reveal more patterns of such activities by filtering out the data of `Coaster Alley` where `Creighton Pavilion-32` is situated to make an initial plot of the situation and make a judgment on the published news. Our visualization approach considered data of every 15 minute's interval as a criterion of sampling and disambiguating dataset prior to visualization. We found that it is raising the `Issue Uncertainty` for structuring, filtering and organizing dataset resulting to `Decision Uncertainty` As shown in Figure 2, three check-in events have been recorded for user id 1102394 within 15 minute's interval into movements table whereas current visualization only visualizes the most recent check-in point. We denote this as `Spatial Determinacy Problem` (Figure 2). Such sampling strategy has raised another issue of missing particular temporal data of an event. We denote this as `Temporal Determinacy Problem` (Figure 3). As shown in Figure 2, no check-in event has been plotted by the high-chart whereas a check-in record has been found into movements table as displayed on spatial selection panel. Both of these determinacy problems raise concerns about `Personal Uncertainty` of the analyst to test the assumptions or hypotheses s/he has been operating with.

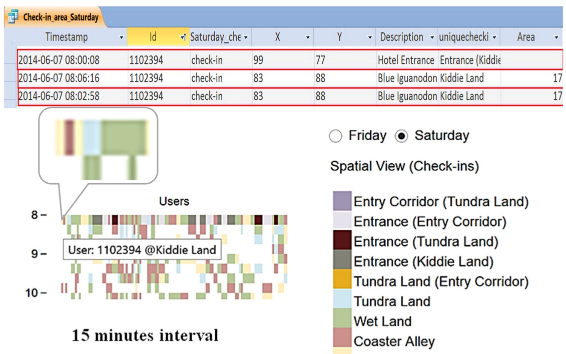

Figure 2: *Spatial Determinacy Problem.*

4.4. Findings on Visualization Biases

The VAST2015 [WCG*15] dataset visualization as shown in Figure 1 shows different patterns of movements, although they are not developed by using any statistical distribution theories i.e, frequency distribution for temporal view [Figure 1(iii)] and spatial distribution for spatial view [Figure 1(iv)]. So, this may create a clustering illusion to analysts leading to cognitive bias while trying to find out a pattern from these plotted data by using $U \times T$ visualization paradigm. Smaller/bigger group activities like group `G6` can be found quite a lot of time through-out the whole visualization. This is a visualization bias where a user is typically unaware of the data values, but is more aware of the position of graphic points from the display.

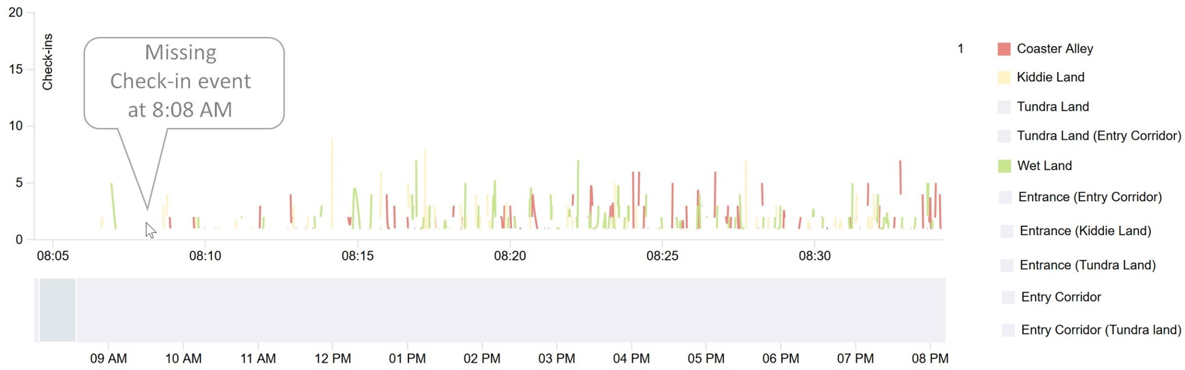

Figure 3: *Temporal Determinacy Problem.*

4.5. Human Trust-Building under Visualization Biases and Uncertainties

By considering the unawareness issues of analysts on errors and limitations of visualizations, we have found that human trust building may get affected due to visualization uncertainties such as spatio-temporal determinacy problems and has got negative impacts on analytic processes due to visualization biases. Muir's [Mui87] description of trust relations between human and machine includes the concept of trust calibration that is influenced by such factors. Analysts have to calibrate their trust not only towards the system but also towards the system outputs, or the findings and insights that have been gained by using the system. The trust in these parts may increase or decrease based on the understanding and awareness of errors or uncertainties that are hidden behind the final system outputs.

5. Provenance for Handling Uncertainties, Biases and Awareness

By making intelligible to analysts the datasets, data configuration and modeling on which their findings are based, provenance techniques can be leveraged to help mitigate uncertainty and distrust between human and machine. On the one hand, data provenance [WXA11], that is information on the types of data that were used as well as details on quality of collected data, enables analysts to track, record and communicate processes [NXW*16] in order to raise awareness of uncertainties. On the other hand, analytic provenance [WXA11], that is the analytic context under which insights were made, enables analysts to review the analysis process or to infer trust levels based on his/her behaviour or interaction with the system. In the following we will describe how these methods can enhance analysis processes that include uncertainties.

5.1. Data Provenance

Uncertainty quantifications for each of the parts within the visual analytics pipeline are the foundation for handling and communicating uncertainties. These uncertainty measures can be propagated and aggregated in order to provide a combined measure that can be related to the system outputs. Furthermore, capturing the process of data transformations and uncertainty information enables the visualization of uncertainties. Finally, provenance techniques enable the exploration of uncertainties and an understanding of how specific data items or dimensions are impacted by different uncertainties.

5.2. Analytic Provenance

Analytic provenance methods for capturing, tracking, managing or organizing evidence [IAX*16] found using a system should be enriched with trust cues about the included uncertainties in order to *Support Uncertainty Aware Sensemaking* [XAJK*15]. Further, trailing human interaction and behaviour might help to infer an analyst's trust level (e.g., which items are of interest or trusted). This information could be leveraged to provide hints about potential problems and biases. Finally, analytic provenance enables the analyst to track and review their analysis as a post-analysis activity in order to detect, assess and mitigate biases.

6. Conclusion

We found from our case study that unawareness of errors and limitations into visualization systems introduce determinacy problems and creates issue uncertainty. Such personal uncertainties of analysts may hinder their decision making process. To make the analysts aware of uncertainties at every stage of data analysis - a background information on how the data were collected or processed (data provenance) and facilities to record, organize revisit their processes (analytic provenance) will aid analysts in this regard.

It is argued that, where uncertainties are fully understood and accounted for in a data analysis, there is greater trust in the acquired knowledge. This notion of trust is perhaps particularly important in crime analysis, where analysts must provide evidence with sufficient clarity and confidence for officials to use in strategic and operational decision-making.

References

[BOL12] Brodlie K., Osorio R. A., Lopes A.: A review of uncertainty in data visualization. 81–109. Expanding the Frontiers of Visual Analytics and Visualization (2012),ISBN 978-1-4471-2803-8. 1

[GD15] GEOFFREY E., DIX A.: Decision making under uncertainty in visualization? IEEE VIS(2015), In: POTTER, Kristi, ed., RÃijdiger WESTERMANN, ed.. VDMU: Workshop on Visualization for Decision Making under Uncertainty. 1, 2

[IAX*16] ISLAM J., ANSLOW C., XU K., WONG B. L. W., ZHANG L.: Towards analytical provenance visualization for criminal intelligence analysis. Computer Graphics and Visual Computing (CGVC) 2016. 4

[LPM01] LEE H., PALMBACH T., MILLER M.: Crime scene handbook. Academic Press, eBook ISBN 9780080507989. 1

[Mui87] MUIR B. M.: Trust between humans and machines, and the design of decision aids. 527–539. International Journal of Man-Machine Studies 27(5-6). 4

[NXW*16] NGUYEN P. H., XU K., WHEAT A., WONG B. L. W., ATTFIELD S., FIELDS B.: Sensepath: understanding the sensemaking process through analytic provenance. 41–50. IEEE Transactions on Visualization and Computer Graphics, 22(1), ISSN 1077-2626. 4

[Ple02] PLEWE B.: The nature of uncertainty in historical geographic information. 431–456. Transactions in GIS, (2002), 6 (4). 3

[SSEK15] STOFFEL F., SACHA D., ELLIS G., KEIM D. A.: Vapd-a visionary system for uncertainty aware decision making in crime analysis. Symposium on Visualization for Decision Making Under Uncertainty at IEEE VIS 2015. 1

[SSK*16] SACHA D., SENARATNE H., KWON B. C., ELLIS G., KEIM D. A.: The role of uncertainty, awareness, and trust in visual analytics. IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology); 22(1):240-9. doi:10.1109/TVCG.2015.2467591. 1, 2

[WCG*15] WHITING M., COOK K., GRINSTEIN G., FALLON J., LIGGETT K., STAHELI D., CROUSER J.: Vast challenge 2015: Mayhem at dinofun world. IEEE Conference on Visual Analytics Science and Technology (VAST), ISBN 978-1-4673-9783-4, DOI: 10.1109/VAST.2015.7347638. 2, 3

[WXA11] WONG B. L. W., XU K., ATTFIELD S.: Provenance for intelligence analysis using visual analytics. CHI 2011: Workshop on Analytic Provenance, 07-08 May 2011, Vancouver, BC, Canada. 4

[XAJK*15] XU K., ATTFIELD S., JANKUN-KELLY T. J., WHEAT A., NGUYEN P. H., SELVARAJ N.: Analytic provenance for sensemaking: a research agenda. 56–64. IEEE Computer Graphics and Applications (2015), 35(3), ISSN 0272-1716. 4