EuroVis Workshop on Visual Analytics (2014)
M. Pohl and J. Roberts (Editors)

Short Papers

From Ill-Defined Problems to Informed Decisions

J.C.Roberts[1], D.Keim[2], T.Hanratty[3], R.R.Rowlingson[4], R.Walker[5], M.Hall[6], Z.Jacobson[7], V.Lavigne[8], C.Rooney[9], M.Varga[10]

[1] Bangor University, [2]University of Konstanz, [3]US Army Research Laboratory, [4]British Telecommunications plc., [5]Middlesex University, [6]EADS, [7]Rogers Communications, [8]Defence Research and Development Canada, [9]Middlesex University, [10]Oxford University.

Abstract

Decision makers such as military leaders and security analysts are increasingly being asked to make decisions on ill-defined problems. These problems may contain uncertain or incomplete data, and are often complex to piece together. Consequently, decision makers rely heavily on intuition, knowledge and experience. We argue for rich narratives that encapsulate both explicit data and implicit knowledge, supported by three levels of provenance: data*,* analytical *and* reasoning*. Our hypotheses is that visual analytics tools and methods can help to provide a valuable means to make sense of these complex data, and to help make this tacit knowledge explicit, to support the construction and presentation of the decision.*

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—: H.5.2 [Information Interfaces and Presentation]: Applications—User Interfaces:

1. Ill-defined problems

In making critical decisions, the problem is often only vaguely defined and the information necessary to make the decisions comes from numerous heterogeneous sources with many uncertainties and complex non-linear interdependencies. In addition, the information is often incomplete and changes over time. This is what in this paper we call an *ill-defined problem*, there is simply not enough information about the problem and the information that is needed to make an informed decision is only partially available [Sch09].

Sometimes, but not always, the problem may be decomposable into sub-problems. Even if problems are decomposable, in many cases the parts are highly interconnected and likely to be non-linearly interdependent. In this case, we cannot divide the problem into smaller chunks to master them individually; instead, we need to find a way to achieve an understanding of the complex problem as a whole.

In the process of making complex decisions, intuition and experience therefore plays an important role. The experience is often conveyed in briefings that are basic narratives. In fact, the hypotheses that are examined in the reasoning process are often communicated as narratives. Narratives can be simple textual descriptions, bullet lists in presentation slides, or annotated hand-drawn diagrams; they are based on the data available and capture potentially important pieces of derived information, reflecting potentially compatible, conflicting, or competing hypotheses that are relevant for the solving the problem. In the reasoning process, the interconnected network of narratives and their properties (e.g., relationships and validity) develop until a decision is derived.

The hypothesis of this paper is that visual analytics tools and methods provide a valuable means for making informed decisions by allowing the user to create rich active narratives, which capture the current understanding of the problem as well as the provenance of the understanding and how it was obtained. Visual representations of the reasoning space, i.e., the developing network of narratives in connection with its underlying data sources, and analytics that enables an understanding of the reasoning process will significantly improve the quality of the reasoning process and the efficiency in reaching informed decisions.

2. Provenance of Data, Analysis and Reasoning

The issue of provenance in visual analytics can be examined at three different conceptual levels. **At the data level** all data have some source, and a path between this source and the system. The data may come from automated systems, network surveillance, from formal intelligence reports, in-

[1] j.c.roberts@bangor.ac.uk, [2]keim@uni-konstanz.de, [3]timothy.p.hanratty.civ@mail.mil, [4]robert.rowlingson@bt.com, [5]rick.t.walker@gmail.com, [6]mark.hall@eads.com, [7]jzj3@rogers.com, [8]valerie.lavigne@drdc-rddc.gc.ca, [9]c.rooney@mdx.ac.uk, [10]margaret.varga@oncology.ox.ac.uk.

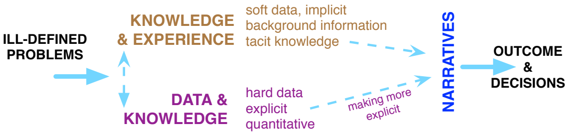

Figure 1: *Hard and soft data is visualized into several narratives. Multiple opinions and uncertainties of the domain can be included to enable better decision making.*

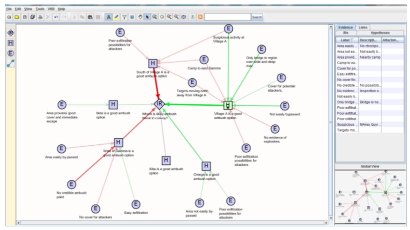

Figure 2: *Network of six alternative hypothesis for ambush analysis; used with permission.*

tercepted communications, or open source text, graphics, or numeric data gleaned from the Internet. Thus provenance, or combination of provenance and data routing, will have at least the potential to impact the nature, quality and reliability of the data. A caution, however: provenance may never be fully perfect and may require 'leaps of faith'. **At the analysis level**: how was the analysis performed? What actions were taken and which techniques were used to process and visualize data? E.g., Visage [RLS*96] captures visualizations, while VisTrails [BCC*05] stores different exploratory actions that a user may have, to enable playback or exploration of different trails. **At the highest level**, reasoning provenance deals with the question of "how did you arrive at these conclusions?". Annotation can be usefully applied to help this situation that can be recalled and shared [GS06, WSD*13].

The issues surrounding data provenance differ by field. In the art world, provenance is concerned with the chain of ownership of a work. In scientific computing, the lineage of output data must be stored – things like parameter settings and code versions – so that results can be reproduced. In intelligence analysis, the sources of data may not be known or visible due to classification levels. Further, data comes with associated reliability issues – things like source uncertainty, or even deliberate misinformation.

The capture of analytical provenance – the actions taken to perform an analysis – is comparatively straightforward within a visual analytic system. Events (key pressed or mouse clicks) or actions (the system level action, such as zoom or filter) can be logged easily. The overall history of interactions can be stored by recording either the state of the application following the interaction, or by recording the action that was performed. Storing states can become inefficient for a large number of states, while storing just interactions gives support for features such as undo/redo.

However, the step from capturing *process* to capturing *reasoning* (stepping from structures to semantic meaning) is much harder to undertake. What is required is to manage the intangible assets of the system. This goes beyond knowledge visualization, which aims to improve the creation and exchange of knowledge by providing richer ways to express what they know [NW05]. But while this externalization can be achieved through think-aloud protocols, the process itself can potentially change the nature of the reasoning that occurs and may reduce task performance [HHA09].

Analysis of such externally captured data is enormously time-consuming. Requiring the analyst to manually annotate their work has similar issues – breaking the cognitive flow of the process risks changing the process itself. Initial work at reconstructing the reasoning process solely from the interactions between the analyst and the system shows promise [DJS*09], and criteria for aiding the effectiveness of this process (semantics of user interactions, information change and degree of interactivity) have been suggested [EB07].

In conducting an analysis, the distinction can be drawn between **hard** and **soft** data (see Figure 1). Hard data is explicit knowledge – typically it's quantitative. Different hard data sets may be combined to use in the analysis process, and here maintaining a provenance trail for the data is important. Hard data has a known source and provenance. Soft data, in contrast, is implicit. In the context of analysis, this means things like general background information, tacit knowledge or personal experience. It is difficult to capture and represent this type of data. However it must be captured if we are able to show reasoning provenance for a situation.

Identifying the difference between hard and soft data is crucial in understanding the importance of provenance. E.g., data fusion – defined as a "multi-level, multifaceted process handling the automatic detection, association, correlation, estimation, and combination of data and information from several sources" [Whi91] – has historically been concerned only with hard data, typically from sensors. But, there is an increasing realization that, in the context of decision support, the issue of soft data cannot be ignored. Decisions can only be understood by considering both soft and hard data in the scenario, which has implications for decision support.

We can store data provenance for hard data, and we can track analytic provenance as the sequence of actions an analyst takes when producing analytic product. Only when we include reasoning provenance deduced from the soft data are we able to complete the audit trail, allowing us to not just understand which data supported the decision making, but also the reasoning that allowed such conclusions to be drawn. The key problem still remains of how can we describe and capture how soft data influences the analyst's actions? If we cannot take the required step from process to intention, and hence to reasoning provenance, then we will never fully understand why certain conclusion were drawn or decisions were made.

3. Displaying narratives and exploring hypotheses

Capturing the provenance data and storing the hard (explicit) data and soft (implicit) data is only part of the challenge. The other is for the user to create and tell the story that encodes the differences and nuances. "Thinking about visualizations in a narrative context can help make them more comprehensible, memorable, and credible to the general public" [MLF*12]. These are narratives that explain the situation and offer insight into possible outcomes or ramifications. Any provenance data that is used in the visualization further adds to the validity and informs the decisions. Therefore, a narrative visualization not only enables the user to understand and to explore different outcomes, but also to tell the *story* to other people.

With Visual Analytics, these narratives end up being explicit representations of the hypothesis. They often include different types of data in their presentation, including explicit information from the original raw dataset, provenance data that shows the processes and manipulation that the data has been subjected to, and implicit information from users' knowledge and experience.

The challenge is to ascertain the best way to display all this ill-defined information. The book "Illuminating the path" [TC05] motivates researchers to investigate tools of analytic reasoning that incorporate note-taking and enable better production, presentation and dissemination. Keim et al. [KKEM10] further this point by encouraging developers to create solutions that document the whole analytic process, keep provenance data and details of findings and discoveries, and help with reporting and storytelling. However, visualizing and manipulating narratives, that enables ill-defined problems to be expressed and manipulated, is not easy.

Narratives are temporal and progressive: they are a sequence of related events that occur over time. The view of stories having a beginning, middle and end is too simplistic. Pragmatically, narratives can be nested, contain other stories, do not need to be expressed linearly, contain several hypothesis (each equally valid and are consequently multivocal opinions), users can question assumptions and conclusions and narratives contain competing hypotheses. E.g., films often use 'flash backs' to fill in the gaps and detail of an event that happened in the past, to give context to some events and provide backstory.

Storytelling and visualization have been linked together for many years, with stories aiding users to remember facts [GP01] and are visual and interactive [LKS13]. Segel and Heer [SH10] review the design space for narrative visualization, they write "An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives". Setel and Heer describe seven styles: magazine, annotated chart, partitioned poster, flow chart, comic strip, slide show and film/video/animation.

Many techniques, presented in the literature, follow the flow-chart design strategy, such as StoryFlow [LWW*13] and CodeTimeline [KS12]. Other tools follow the comic strip approach (where frames are positioned side-by-side) or slide-show (where the next slide replaces the previous), this is exemplified by storyboarding techniques (such as) Walker et al. [WaCP*13] who use it to display microblog data. The techniques encompassed by storyboarding, especially the transition of frames and different locations are heading towards the design principle of a *grand tour*. E.g., The Excel^{TM} extension GeoFlow, enables users to display geographical information spatially, annotate and to tour from difference key locations in 3D; while Lee et al. [LKS13] control the slide-show type visualizations through gestures.

Stories are often told or explored together. Computer Supported Cooperative Work (CSCW) enables everything to be shared and edited by teams. This may have security and ownership concerns, and can give rise to alternative viewpoints and arguments between collaborating users.

Arguments have been visualized since the Wigmore maps (1931), Toulmin's argument scheme in 1958 and Horn's argument maps of debates in the 1990's [KBSC03]. In later years Computer Supported Argument Visualization (CSAV) has become an established genre of visualization, achieving the display of arguments through automated tools, see Kirschner et al. [KBSC03]. These arguments often represent multiple-views [Rob07]. In fact, these are alternative viewpoints, each viewpoint is an opinion of an expert and is just as valid as the other. E.g., in security analysis, ambush locations of the obvious and the less obvious alternatives can be considered. This is shown in Figure 2 where a hypothesis is introduced that suggests that south of Village A (shown in green) is a good ambush option [VA10]). The Analyst enters properties of the hypothesis based on his intuition and his information sources, which can be rated as being completely, usually, fairly reliable etc. [HIE06].

To illustrate our ideas, we provide two case-studies (cyber security and human terrain analysis) that illustrate the challenges of ill-defined problems.

4. Case Study 1: Cyber Security

Cyber security, such as a 24x7 security operations center (SOC), serves as an ideal case study. The ill-defined problem is to assess whether the weak signals of a potential cyber attack are sufficient threats to the business to require some action. This operational impact on the organization is continuously assessed by aiming to maintain a situational awareness of the state of the environment. This situational awareness is continuously evolving, and difficult to achieve since data, knowledge and intelligence are from numerous heterogeneous sources, with many uncertainties and complex nonlinear interdependencies. The narratives, hypotheses, provenance, etc. all contribute to achieving this desired state of situational awareness. This transformation of hard and soft data into meaningful information in context, then ultimately actionable knowledge, is vital to make informed decisions.

Businesses are frequently exposed to novel attacks launched by highly skilled adversaries. Visual analytics has been shown to be a useful tool in a cyber SOC in supporting the assessment and the decision-making process in responding actions [RHS*13]. Although analysts typically operate as part of a team, visual analytics tools are generally single user. It is critical that the skills and capabilities of the defenders, the analysts in the SOC, are fully utilized. The need for teamwork demands support for communicating and assessing the interconnected network of narratives that evolve in cyber security operations; building on:

- Implicit knowledge – including the knowledge of attack techniques, an awareness of new malware, understanding of typical network behaviors, understanding of traffic patterns typical to the business, assessment of the potential impact on the business of a successful attack.
- Explicit knowledge and data – feeds from various tools and sensors on the network and hosts, network topology and device parameters. This knowledge and data will be used in the analytic process to feed the provenance of informed decisions. Elements of a case study could cover:
- Automated provenance capture;
- Semi-automated knowledge/ hypothesis/ problem statement/ solution capture;
- Traceability, audit-ability and knowledge capture through logging.

The challenge is to capture how a team moves from *problem* to *solution*, to understand its use of explicit and implicit knowledge, and to be able to reuse that process in related attack scenarios.

5. Case Study 2: Human Terrain Analysis

Characterized by a multifaceted, multidisciplinary approach, human terrain analysis (HTA) aims to describe and predict geospatial and temporal patterns of human behavior by analyzing the attributes, actions, reactions and interactions of groups or individuals in the context of their environment [NGI08]. For the military, effective HTA is critical to combating the continual commitment to Counter-insurgency (COIN); Stability, Security, Transition, and Reconstruction (SSTR); and Humanitarian Assistance and Disaster Relief (HADR) operations. To emphasize just how important HTA is to current military operating procedures, General Petraeus (US Army Central Command Commander) said, "You have to understand not just what we call the military terrain the high ground and low ground. It's about understanding the human terrain, really understanding it [Deh08]."

Challenging the HTA process are numerous factors, not the least of which are the complexity of the questions being proffered. HTA by its very nature is framed by ill-defined problems focused on understanding concepts like relationships, sentiment, trends, activities and events. In addition, with HTA, greater reliance is put on inference rather that direct observation, so uncertainties are also increased. Further confounding this domain is the abundance of non-traditional information sources. One such source that has escalated in importance over the past decade is open source information. Defined in its simplest terms, open source information is publicly available information appearing in print or electronic form [USA06]. Electronic open source information (EOI), specifically news feeds, blogs and other social media, provides a unique opportunity to collect and evaluate salient topics, trends and sentiments within a military area of interest. Interpreted correctly, EOI can also provide valuable insights in determining opinions, values, cultural nuances and other sociopilitical aspects [USA06].

While point solutions to this domain are being developed, what is lacking are composite solutions that effectively and efficiently provide the ability to iteratively build rich narratives that capture the multifaceted dimensions associated with HTA. Critical to this iterative process are visual analytics solutions that facilitate synthesizing data to decisions. Capitalizing on the human capacity for spatial reasoning, visual analytics enhance the decision maker's understanding of the underlying decision space by augmenting the assimilation of complex relationships [HHMR13]. That said, the key to successful HTA will be found in drawing upon the strengths of decision makers to visually interpret human terrain data while incrementally exploring complementary and contradictory narratives about the decision space gaining valuable insights in determining opinions, values, cultural nuances and other sociopolitical aspects within a given military area of interest.

6. Summary & Conclusions

So far, we have visual analytics tools that provide provenance, enable narratives to be expressed in frames (such as storyboards), provide schematic diagrams to display arguments, and enable the different viewpoints of experts to be displayed together in multiple coordinated views or side-by-side windows. However, no system fully enables or supports each of these aspects together. They are currently individual tools that operate on their own and do not capture or visualize the tacit, soft or implicit knowledge.

To fully support ill-defined questions, we need systems that visualize hypotheses, which develop over time; that change because of the arrival of 'new information' or the application of a 'new process', methods to visualize competing hypothesis or complementary theories (that would support and enhance the strength of a particular argument), each depicting different degrees of certainty. While many pieces of the puzzle exist, there is still much research to be performed to aid decisions makers to better turn ill-defined questions to informed decisions, through the use of visual analytics.

7. Acknowledgements

This work was initiated through a collaboration formed at the NATO Science & Technology community IST-116-RSY-028 Symposium on Visual Analytics 27-28 Oct 2013, and IST-117-RWS-017 Workshop on Visualization for Analysis. The authors acknowledge the NATO S&T community and in particular Richard May for his encouragement and advice.

References

[BCC*05] BAVOIL L., CALLAHAN S., CROSSNO P., FREIRE J., SCHEIDEGGER C., SILVA C., VO H.: Vistrails: enabling interactive multiple-view visualizations. In *Visualization, 2005. VIS 05. IEEE* (Oct 2005), pp. 135–142. doi:10.1109/VISUAL.2005.1532788.

[Deh08] DEHGHANPISHEH B.: Scions of the Surge. Newsweek, March 2008.

[DJS*09] DOU W., JEONG D. H., STUKES F., RIBARSKY W., LIPFORD H., CHANG R.: Recovering reasoning processes from user interactions. *Computer Graphics and Applications, IEEE 29*, 3 (May 2009), 52–61. doi:10.1109/MCG.2009.49.

[EB07] EPPLER M. J., BURKHARD R. A.: Visual representations in knowledge management: framework and cases. *Journal of Knowledge Management 11*, 4 (2007), 112–122. doi:10.1108/13673270710762756.

[GP01] GERSHON N., PAGE W.: What storytelling can do for information visualization. *Commun. ACM 44*, 8 (Aug. 2001), 31–37. doi:10.1145/381641.381653.

[GS06] GROTH D., STREEFKERK K.: Provenance and annotation for visual exploration systems. *Visualization and Computer Graphics, IEEE Transactions on 12*, 6 (Nov 2006), 1500–1510. doi:10.1109/TVCG.2006.101.

[HHA09] HERTZUM M., HANSEN K. D., ANDERSEN H. H. K.: Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behav. Inf. Technol. 28*, 2 (Mar. 2009), 165–181. doi:10.1080/01449290701773842.

[HHMR13] HANRATTY T., HEILMAN E., MITTRICK M., RICHARDSON J.: A visual analytic for improving human terrain understanding. In *18th International Command and Control Research and Technology Symposium (ICCRT)* (Arlington, VA., June 2013).

[HIE06] Human intelligence collector operations. FM 2-22.3 (FM 34-52). US Department of the Army, September 2006.

[KBSC03] KIRSCHNER P. A., BUCKINGHAM-SHUM S. J., CARR C. S.: *Visualizing argumentation: Software tools for collaborative and educational sense-making*. Springer, 2003.

[KKEM10] KEIM D. A., KOHLHAMMER J., ELLIS G., MANSMANN F.: *Mastering The Information Age-Solving Problems with Visual Analytics*. Eurographics Association, 2010.

[KS12] KUHN A., STOCKER M.: Codetimeline: Storytelling with versioning data. In *Software Engineering (ICSE), 2012 34th International Conference on* (June 2012), pp. 1333–1336. doi:10.1109/ICSE.2012.6227086.

[LKS13] LEE B., KAZI R., SMITH G.: Sketchstory: Telling more engaging stories with data through freeform sketching. *Visualization and Computer Graphics, IEEE Transactions on 19*, 12 (Dec 2013), 2416–2425. doi:10.1109/TVCG.2013.191.

[LWW*13] LIU S., WU Y., WEI E., LIU M., LIU Y.: Storyflow: Tracking the evolution of stories. *Visualization and Computer Graphics, IEEE Transactions on 19*, 12 (Dec 2013), 2436–2445. doi:10.1109/TVCG.2013.196.

[MLF*12] MA K.-L., LIAO I., FRAZIER J., HAUSER H., KOSTIS H.-N.: Scientific storytelling using visualization. *Computer Graphics and Applications, IEEE 32*, 1 (Jan 2012), 12–19. doi:10.1109/MCG.2012.24.

[NGI08] National Geospatial-Intelligence Agency (NGIA) briefing at the US Geospatial Intelligence Foundation. GEOINT 2008 Symposium, Nashville, Tenn, October 2008.

[NW05] NOVAK J., WURST M.: Knowledge and information visualization. Springer-Verlag, Berlin, Heidelberg, 2005, ch. Collaborative Knowledge Visualization for Cross-community Learning, pp. 95–116. doi:10.1007/11510154_6.

[RHS*13] ROWLINGSON R. R., HEALING A., SHITTU R., MATTHEWS S., GHANEA-HERCOCK R.: Visual analytics in the cyber security operations centre. In *NATO IST-116 Symposium on Visual Analytics*. Defence Academy, Shrivenham, UK, October 2013.

[RLS*96] ROTH S. F., LUCAS P., SENN J. A., GOMBERG C. C., BURKS M. B., STROFFOLINO P. J., KOLOJECHICK A., DUNMIRE C.: Visage: a user interface environment for exploring information. In *Information Visualization '96, Proceedings IEEE Symposium on* (1996), IEEE, pp. 3–12,116. doi:10.1109/INFVIS.1996.559210.

[Rob07] ROBERTS J. C.: State of the art: Coordinated multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV '07. Fifth International Conference on* (July 2007), pp. 61–71. doi:10.1109/CMV.2007.20.

[Sch09] SCHACTER D.: *Psychology, Second Edition*. New York: Worth Publishers, 2009.

[SH10] SEGEL E., HEER J.: Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on 16*, 6 (Nov 2010), 1139–1148. doi:10.1109/TVCG.2010.179.

[TC05] THOMAS J. J., COOK K. A.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[USA06] US Army Field Manual (FM) 2-22.3. Human Intelligence Collection Operations, US Army, September 2006.

[VA10] VARGA M. J., ADAMS K.: Interactive hypothesis visualization. In *NATO Workshop on Visualising Networks: Coping with Change and Uncertainty* (October 2010).

[WaCP*13] WALKER R., AP CENYDD L., POP S., MILES H. C., HUGHES C. J., TEAHAN W. J., ROBERTS J. C.: Storyboarding for visual analytics. *Information Visualization* (2013). doi:10.1177/1473871613487089.

[Whi91] WHITE F. E.: *Data fusion lexicon*. Tech. rep., DTIC Document, 1991.

[WSD*13] WALKER R., SLINGSBY A., DYKES J., XU K., WOOD J., NGUYEN P., STEPHENS D., WONG B., ZHENG Y.: An extensible framework for provenance in human terrain visual analytics. *Visualization and Computer Graphics, IEEE Transactions on 19*, 12 (Dec 2013), 2139–2148. doi:10.1109/TVCG.2013.132.