Towards Smart Visual Data Exploration Tools

Lars Grammel and Margaret-Anne Storey

University of Victoria

Abstract

The semantic information made available by the Smart Internet could enable visual data exploration tools to provide users with a personalized and task-centric data analysis experience. We describe some of the potential opportunities and challenges of leveraging user profiles, task context, and semantic meta-data to improve visual data exploration and analysis.

1 Motivation

Visual data exploration and analysis is a time-consuming activity that requires both domain specific knowledge (about the data being explored) and analytical expertise (in both visualizations and data analysis). It is often performed in the context of a higher-level task. When visual data exploration is carried out by a single analyst, this process would ideally be tailored to the individual, to the data being explored, and to the contextual task. The Smart Internet [3], "a platform for automatic, dynamic aggregation of data and services for the purpose of supporting each user's goals, tasks and concerns, both cognitively and socially" [2], could provide such semantic information that would enable streamlining this activity and that would make visual data exploration tools accessible to a larger audience.

In this position statement, we describe ideas of how user profiles, task context, and semantic meta-data could potentially be employed to create adaptive visual data exploration tools, and some of the challenges that still need to be overcome[1]. Smart visual data exploration tools that adopt to the user and his/her task could be used as components of the Personal Web [2] that provide data analysis capabilities as required by the user.

We will first briefly summarize the gap that exist between generic visual data exploration tools and custom designed visualizations, and examine at related research that addresses this gap. Next, we will describe potential opportunities and challenges of leveraging user profiles, task context and semantic meta-data. Finally, we describe Choosel, a visual data exploration tool we have developed, and the adoption capabilities we are planning to add to it.

2 Related Work

Visual data exploration tools allow users to analyze and gain insights into data sets visually. Commercial tools such as IBM Cognos 8 Business Intelligence[2], Tableau[3], and Tibco Spotfire[4] provide standard visualization and analysis tools that work with diverse sets of data. These differ from custom-designed, data-specific visualizations in that they can be applied out of the box (eliminating the delay required for developing a domain-specific visualization). However, the resulting visualizations are inherently oblivious of the data set, the user and the task context and thus

[1] The discussion is limited to desktop computers, ignoring device-specific adaption.
[2] http://www-01.ibm.com/software/data/cognos/products/cognos-8-business-intelligence/
[3] http://www.tableausoftware.com/
[4] http://spotfire.tibco.com/

are potentially less effective than custom made visualizations (which are designed for a specific data set and for a typical user group and task).

This gap between visual data exploration tools and custom designed visualizations is partially addressed by automatic visualization tools. Automatic visualization is concerned with creating visualizations without the involvement of a visualization designer. Automatic visualization tools such as APT [9] usually take a data set and a list of data properties as input. Some tools attempt to leverage task descriptions [1, 11], however, tasks are usually hard to describe and the visualization intent may be difficult to elicit [10]. Automatic visualization functionality has also been integrated into the user interface (UI) of visual data exploration tools such as Tableau [10]. Gilson *et al.* employed semantic domain descriptions for automatic visualization using domain, bridging, and visualization ontologies [4]. The domain ontology described the problem domain, the visualization ontology described a specific visualization, and the bridging ontology contained expert knowledge on how the two should be mapped. Despite the progress researchers made in automatic visualization, leveraging task context, user profile and semantic meta-data is still require further investigation.

3 Towards Smart Visual Data Exploration Tools

In this section, we describe how a hypothetical visual data exploration tool could adopt itself based on user profiles, task context and semantic meta-data, and what challenges we foresee. In our scenario, the data exploration tool would rely on the Smart Internet to provide the user profile and task context. A data warehouse would provide the data as well as semantic meta-data.

3.1 User Profile

Four adaption methods that could be derived from the user profile are preference-based adaption, impairment-based adaption, skill-based adaption, and social-network-based adaption.

Preference-based adaption is based on user preferences, such as language settings or regional preferences (e.g., time zone or currency). Impairment-based adaption takes the impairments of a specific user into account (e.g. color deficiencies). Skill-based adaption takes the user's skill level into account to provide support and hide/offer functionality. Social-network based adaption considers potential support in the social network during UI customization.

The following is an example of using skill-based and social-network based adaption. User Carl is analyzing some business data of interest. He selects several data attributes he wants to analyze. Because he is inexperienced in creating visualizations, the system decides to present the most appropriate visualization, instead of letting him choose from a range of potential visualizations. While Carl tries to analyze the visualization, the system detects that Carl performs repetitive operations and seems to be stuck. Depending on the level of expertise and availability of people in Carl's company's social network, the system could either recommend a local expert for Carl to contact, or display a help page on how to interpret such a visualization.

Challenges to employing user profile-based adaption methods include how to determine the skill level of a user, detecting skill level changes, knowing when a user is stuck, and how to support the user in a non-intrusive manner.

3.2 Task Context

The task context models the high-level task that the user is working on. However, there are several levels how tasks and activities can be modeled and are connected [5]. While the Smart Internet might be able to provide information about the higher level task, knowing concrete intentions for using visualizations at a lower task level might still be challenging [10].

The task context can be used to select relevant data from the data warehouse, to guide automatic visualization algorithms, and to find visualizations used previously by others on similar tasks. Challenges are the modeling of the task, and how it can be connected to automatic visualization algorithms.

3.3 Semantic Meta-Data

Semantic meta-data describes how data and data attributes in the data warehouse relate to real-world processes and concepts. This includes relationships to concepts such as units, relationships to other data attributes, and formulas how data values are calculated. For example, semantic meta-data could specify that the budget sales data attribute refers to a projection for future sales in US dollars. The general concept of sales could then link to other data attributes with the same underlying type, such as the real sales, or to other sales estimations.

Similar to the task context, semantic meta-data can be used to direct access to parts of the data warehouse and to inform automatic visualization algorithms [7]. It can also be used to suggest other possible data attributes to visualize or to explain observed relationships in the data.

For example, when the user requests to visualize actual and estimated sales, the automatic visualization algorithm could take into account the relationship between those two data attributes. It could then choose a visualization that highlights the differences between actual and estimated sales for past months.

We are planning to explore different ways of leveraging such semantic meta-data as part of our development of the visual data exploration tool Choosel.

4 Choosel

We have started developing a visual data exploration tool called Choosel[5] that will enable us to investigate some of these concepts [6]. Choosel is a web-based tool that aims at facilitating flexible visual data exploration for information visualization novices (see Figure 1). It supports the iterative construction of multiple coordinated views during the visual data analysis process.

We have built a work item exploration tool (Work Item Explorer) and a bio-medical data exploration tool (Bio-Mixer[6]), which are based on the Choosel framework. The Work Item Explorer is an example of our vision of software mashups [8] and allows developers to analyze work item data in different visualization such as graphs, time lines, and charts. An online version of Choosel is also available to enable web users to explore any kind of supplied tabular data[7].

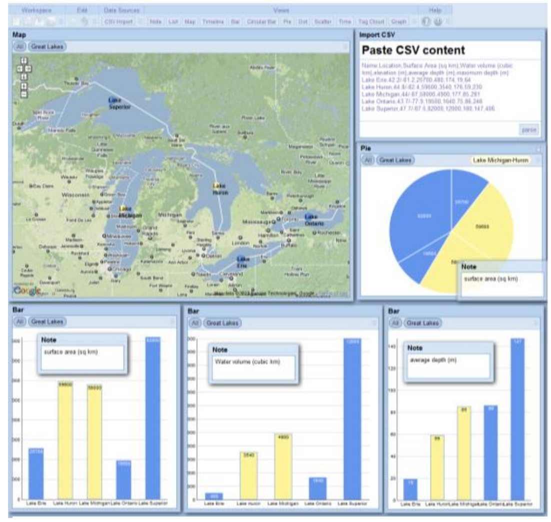

Figure 1: Choosel Visual Data Exploration Environment

We are currently working on adding automatic visualization and visualization recommendation to Choosel. As part of this, we are specifically looking at using data sets that are annotated with semantic meta-data.

5 Conclusion

Smart visual data exploration tools have the potential to make visual data exploration more effective and accessible for inexperienced users. Leveraging user profiles, task context and semantic meta-data are fruitful avenues for adopting such tools and directing automatic visualization tools. We believe that smart visual data exploration tools will be an important part of the Smart Internet and enable users to rapidly understand relevant data.

[5] http://code.google.com/p/choosel/
[6] http://bio-mixer.appspot.com/
[7] http://choosel-mashups.appspot.com/

Acknowledgements

The authors would like to thank Chris Bennett for his editing support. This research was funded by an IBM CAS PhD Fellowship.

References

[1] Stephen M. Casner. Task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics (TOG)*, 10(2):111–151, 1991.

[2] Mark Chignell, James R. Cordy, Joanna W. Ng, and Yelena Yesha. First symposium on the personal web. In *CASCON '10: Proceedings of the 2010 Conference of the Center for Advanced Studies on Collaborative Research*, 2010.

[3] Mark Chignell, James R. Cordy, Joanna W. Ng, and Yelena Yesha, editors. *The Smart Internet*, volume 6400 of *Lecture Notes in Computer Science*. Springer, 1st edition, 2010.

[4] Owen Gilson, Nuno Silva, Phil W. Grant, and Min Chen. From web data to visualization via ontology mapping. In *Computer Graphics Forum*, volume 27, pages 959–966. Blackwell Publishing Ltd, 2008.

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

[6] Lars Grammel and Margaret-Anne Storey. Poster: Choosel - web-based visualization construction and coordination for information visualization novices. In *InfoVis 2010: Proceedings of the IEEE Symposium on Information Visualization*. IEEE Computer Society.

[7] Lars Grammel, Melanie Tory, and Margaret-Anne Storey. How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics (TVCG, Proc. InfoVis 2010)*, 16(6), 2010.

[8] Lars Grammel, Christoph Treude, and Margaret-Anne Storey. Mashup environments in software engineering. In *Web2SE '10: Proceedings of the 1st Workshop on Web 2.0 for Software Engineering*, pages 24–25, New York, NY, USA, 2010. ACM.

[9] Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.

[10] Jock Mackinlay, Pat Hanrahan, and Chris Stolte. Show Me: Automatic presentation for visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1137–1144, 2007.

[11] Steven F. Roth and Joe Mattis. Automating the presentation of information. In *Seventh IEEE Conference on Artificial Intelligence Applications, 1991.*, 1991.