Articulate2: Toward a Conversational Interface for Visual Data Exploration

Jillian Aurisano*
Electronic Visualization Laboratory
University of Illinois at Chicago

Abhinav Kumar†
University of Illinois at Chicago

Alberto Gonzalez‡
University of Hawaii at Manoa

Jason Leigh§
University of Hawaii at Manoa

Barbara Di Eugenio¶
University of Illinois at Chicago

Andrew Johnson‖
Electronic Visualization Laboratory
University of Illinois at Chicago

ABSTRACT

'InfoVis novices' have been found to struggle with visual data exploration. A 'conversational interface' which would take natural language inputs to visualization generation and modification, while maintaining a history of the requests, visualizations and findings of the user, has the potential to ameliorate many of these challenges. We present Articulate2, initial work toward a conversational interface to visual data exploration.

1 INTRODUCTION

Visual data exploration is an increasingly essential activity in many data intensive fields. In response, a variety of tools have been developed to allow InfoVis novices to visualize and explore their data, such as Tableau and ManyEyes [15]. However, even with the aid of these tools, such users may face more fundamental challenges with visualization construction and sensemaking.

Recent research has suggested that vis novices make a variety of errors in selecting visual templates or mapping data variables to visual encodings, hampering visualization creation. In addition, users struggle to translate high-level questions and complex analysis tasks into sequences of appropriate visual representations [7] and [9].

Natural language inputs to visualization have the potential to aid users in visualization construction, by allowing users to directly ask high-level questions about data, without translating these questions into low level interactions with a graphical interface. Our initial work in developing a natural language interface in [13], has produced promising results, suggesting the benefit of this kind of input for data visualization. Related promising work can be found in [4] which handles ambiguity in natural language inputs for visualization.

However, visual data analysis involves repeated cycles of question formation, visualization construction and refinement, and insight generation. While many tools have been created to aid in this sensemaking process [3, 10], it is challenging to infer high-level sensemaking tasks from the low-level interaction primitives [6, 5].

To respond to this challenge, we present initial work toward building a **conversational interface** which will allow users to engage in an extended dialog with a visualization system to analyze data. Our aim is to enable users to pose high-level questions, reference prior visualizations and record their findings through spoken natural language inputs and gestures. Using analysis results from our exploratory observational study, presented in [1], we have developed a prototype, Articulate2, which accepts spoken requests to the system, classifies these requests into major types, parses into a logical form to produce visualization specifications for bar charts, line charts and geospatial heat maps which are then presented to the user on a large and flexible display canvas. Ongoing work will expand this system to handle references to prior states and capture and visualize user sensemaking.

*e-mail: jauris2@uic.edu
†e-mail:akumar34@uic.edu
‡e-mail:agon@hawaii.edu
§e-mail:leighj@hawaii.edu
¶e-mail:bdieugen@uic.edu
‖e-mail: ajohnson@uic.edu

2 DATA COLLECTION

To collect data about how users interact with a conversational interface in visual data analysis, we performed a 'Wizard of Oz' style study in which a user interacted with a remote 'data analysis expert' (DAE). Users performed an analysis of city of Chicago crime data between 2010-2014 by asking questions of the DAE. Users were encouraged to speak directly and naturally to the DAE and use gestures when necessary to convey their question. Users were also encouraged to think-aloud to describe their observations and reasoning.

The data analysis expert had access to two live video feeds of the user as well as a mirror of the user's display. The DAE could communicate with the user through a chat box and a status bar, simulating a conversational interface. The DAE used Tableau to generative visualizations, which were then shared with the user using Sage2 [11], a collaborative large-display middleware. Visualizations on screen could be moved and closed by the DAE in response to user requests.

15 subjects, 8 male and 7 female between the ages of 18 and 34 participated in the study, which lasted from 45 minute to an hour.

Video recordings of the sessions were saved and transcribed. A team of three coders developed a coding scheme through multiple passes over transcripts and an iterative refinement of codes. User utterances were first divided into two categories: actionable utterances for which a system response is expected, such as a visualization or chat response, and non-actionable utterances, generally when the user was thinking aloud. Actionable commands were then divided into six categories: 1) request for a new visualization, 2)modification to an existing visualization, 3) fact-based question, 4) request for window management, such as closing or repositioning a visualization, 5) clarification requests and 6) expression of preference.

Initial analysis has indicated that 85 percent of user utterances are not directly actionable, but rather provide context in determining a valuable response and indicate user sensemaking tasks. The remaining fifteen percent of user utterances are broken down as follows: 1) request for a new visualization (12 percent), 2) modification to an existing visualization (30 percent), 3) fact-based question

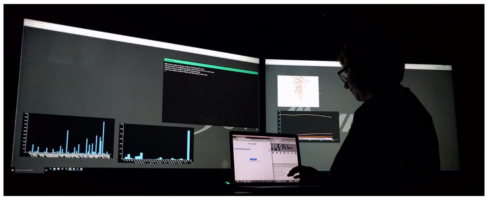

Figure 1: User interacting with Articulate2 through natural language inputs to analyze city of Chicago crime data.

(12 percent), 4) request for window management (13 percent), 5) clarification requests (18 percent) and 6) expression of preference (18 percent).

3 System architecture

Our prototype system recieves spoken inputs from the user and processes these into visualization specifications that can be interpreted by a visualization executer and rendered to the user.

Spoken inputs are recieved through an input service, an html page which parses speech into text using Google's webspeech API. The spoken inputs are passed to a 'smart-hub' java web service which classifies the request into one of the six actionable types described in the previous section. Apache OpenNLP [12] was used to generate unigrams, bigrams, trigrams, chunking, and tagged unigrams, while Stanford Parsers implemented Collins rules [2] were used to obtain the headword. The feature vector is comprised of 7,244 total features. We used Weka [8] to experiment with several classifiers. Currently SVM performs best at 87.65 percent accuracy.

For utterances classified as visualization requests, a logical form is obtained which is used to develop an SQL query as well as a description of the axes and chart type, which is used to form the visualization specification. Utterances for window management use a keyword extraction approach to determine whether the command relates to closing or repositioning a visualization. At present, only requests to close the most recent visualization are supported.

The visualization specification is passed to the Articulate2 Visualization Manager, which runs within Sage2 [11] a collaborative middlewear built through html, javascript and node.js, which scales from personal displays to tiled-display walls. Visualization specifications are parsed and rendered on a flexible canvas using vega.js [14]. The Articulate2 Visualization Manager also displays the user requests in a dialog box, and stores references to the visualizations produced, allowing for modification or repositioning requests. Visualizations may be repositioned or closed by the user using a pointer or using the Sage2 UI described in [11].

4 Results

A controlled study cannot be performed until the components of the system are completed. However, our inspection of the pipeline results for new visualization requests are promising. Some example queries that the system can process include the following sequence, concerning crimes in two Chicago neighborhoods 'the Loop' and 'River North', as well as two data attributes 'crime type' (eg. theft, assault, burglary) and 'location type' (eg. street, sidewalk, parking lot). 1) "Show me a map of crimes in River North and the Loop." 2) "Can you show it around the Loop by year broken down by crime type?" 3) "Can I see assaults in the Loop by location type?" 4) "Can you close the graph?" A user interacting with the system is shown in Figure 1.

5 Future Work

Future work will focus in three major directions. First, we will analyze recordings and transcripts from our observational study, to understand how users reference prior visualizations or visual objects in formulating new questions. Then, we will prototype approaches to capture references in spoken queries, integrating speech and gesture. Second, we will examine approaches to integrate and visualize user sensemaking tasks in visual data analysis. This will involve first an analysis of non-actionable utterances from our study, to understand how users express their sensemaking process through a conversational interface. Then, we will test ways to capture and respond to spoken expressions of user sensemaking activities, as well as how to use this information in responding to visualization requests. Finally, we aim to bring these components together into a functional conversational prototype, using a natural language dialog model and stored visualization and utterance histories.

Acknowledgements

This work is supported by NSF grant IIS-1445751: EAGER: Collaborative Research: Articulate: Augmenting Data Visualization With Natural Language Interaction. SAGE2 is a trademark of the University of Illinois Board of Trustees.

References

[1] J. Aurisano, A. Kumar, A. Gonzales, K. Reda, J. Leigh, B. DiEugenio, and A. Johnson. Show me data. observational study of a conversational interface in visual data exploration, 2015. Poster presented at the Information Visualization conference at IEEE VisWeek in Chicago, IL in October 2015.

[2] M. Collins. Head-driven statistical models for natural language parsing. *Computational Linguistics*, 29(4):589–637, 2003.

[3] K. Cook, N. Cramer, D. Israel, M. Wolverton, J. Bruce, R. Burtner, and A. Endert. Mixed-initiative visual analytics using task-driven recommendations. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*, pages 9–16. IEEE, 2015.

[4] T. Gao, M. Dontcheva, E. Adar, Z. Liu, and K. G. Karahalios. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 489–500. ACM, 2015.

[5] D. Gotz and Z. Wen. Behavior-driven visualization recommendation. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 315–324. ACM, 2009.

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

[7] L. Grammel, M. Tory, and M. A. Storey. How information visualization novices construct visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):943–952, 2010.

[8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[9] J. Heer, F. Van Ham, S. Carpendale, C. Weaver, and P. Isenberg. Creation and collaboration: Engaging new audiences for information visualization. In *Information Visualization*, pages 92–133. Springer, 2008.

[10] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE transactions on visualization and computer graphics*, 20(12):1633–1642, 2014.

[11] T. Marrinan, J. Aurisano, A. Nishimoto, K. Bharadwaj, V. Mateevitsi, L. Renambot, L. Long, A. Johnson, and J. Leigh. Sage2: A new approach for data intensive collaboration using scalable resolution shared displays. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, pages 177–186. IEEE, 2014.

[12] A. OpenNLP. Apache software foundation. *URL http://opennlp.apache.org*, 2011.

[13] Y. Sun, J. Leigh, A. Johnson, and S. Lee. Articulate: A semi-automated model for translating natural language queries into meaningful visualizations. In *Smart Graphics*, pages 184–195. Springer, 2010.

[14] Trifacta. Vega: A Visualization Grammar, 2014.

[15] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1121–1128, 2007.