Poster Abstracts at Eurographics/ IEEE-VGTC Symposium on Visualization (2010)
D. Auber, G. Melançon, T. Munzner, and D. Weiskopf (Chairs)

Experimental Evaluation of a Stacked Graph Visualization

Alejandro Toledo, Kingkran Sookhanapibarn, Ruck Thawonmas, and Frank Rinaldo

Ritsumeikan University, Japan

Abstract

We introduce an experimental evaluation of a Stacked Graph Visualization (SGV) that was designed to facilitate the production of views required for different visual analytics tasks. Our interface encourages users to gain insight into a data set of 510 time series of 255 kinds of jobs. We include a case study that applies the SGV in a visual analytics environment and a usability study that evaluates the correlation between three implicit interest indicators (dwell time, view revisits, and mouse activity) and the explicit rating for a single view. The SGV was developed to record the views produced by 14 users and the explicit rating of those views. Using the data collected by the SGV, the individual ratings and some combinations of explicit ratings were analyzed and compared with the explicit rating. We found that the three implicit interest indicators had a strong correlation with explicit interest.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

Stacked graph visualizations have been widely promoted as an approach for time series visualization ([WK06], [HVW09]). The stacked graph is a method for visualizing changes in a set of items, where the sum of their values is as important as the individual items. One important issue associated with the SGV is that viewers, from simple tasks, can quickly gain insight into the data set with a glimpse into the visualization as a whole. A comprehensive analytical task plan, however, requires trying several analytical paths in the search for a solution, due to the multiple perspectives on the data from the SGV's usage. The key idea of this paper is to conduct an experimental evaluation of the SGV's usage for measuring the influence of three implicit interest indicators: *dwell time*, *view revisits*, and *mouse activity* in the search for a solution of two questions. The approach we adopt makes use of implicit/explicit interest indicators from recommender systems ([CLWB01], [RS01]) and focus on the analysis of these indicators using a Kruskal-Wallis test.

2. Usability Evaluation Procedures

We conducted a usability evaluation of the SGV with participants from our laboratory. For the usability study, we asked participants to answer two different kinds of questions by using the features of our visualization tool. Our general hypothesis was that participants would produce more *dwell time*, *revisits*, and *mouse activity* on views they explicitly rate as useful.

Fourteen participants participated in the usability evaluation. The participants were graduate students, who had no prior knowledge of and experience with the SGV. Participants took part in individual sessions that lasted an average of 40 minutes. At the beginning of each session, participants were given a brief tutorial on the use of the SGV. Furthermore, they were allowed to freely use our tool during five minutes so that they were accustumed with the SGV before answering the questions. Two questions were then given to each participant one by one. According to the visual tasks characterization (*identify*, *compare*, *correlate*) [ZF98], the questions were as follows:

1. *From 1850 to 1950, what is the most popular male occupation? (**a.** Operative, **b.** Military, **c.** Farmer, **d.** Laborer)*
2. *Which three of the following are true?*
   - **a.** *There are four types of male occupations starting with the prefix "Mil"*

/

b. *In 2000, Military (male) was less popular than Teacher (male)*
c. *The highest peak of Secretary (male) has a greater occurrence ratio than the highest peak of Farmer (female).*
d. *Household Worker (female) and Operative (female) have a similar trend*
e. *The drop in the popularity of Farmer (male) precedes an increase in the popularity of Farm Laborer (male)*

For each question, participants were given a time limit of 4 minutes to complete the question. Each participant was timed, observations were gathered, and the participant's answers was collected for each question.

After answering each question, participants were asked to rate the usefulness of the views he/she produced. As shown in Figure 1, our system provides participants with an evaluation interface containing, on the left, the views produced, and on the right, a number of boxes grouped as HIGH (very useful), MIDDLE, and LOW (not useful). According to this criteria, participants were to move their views from the left area to the right area.

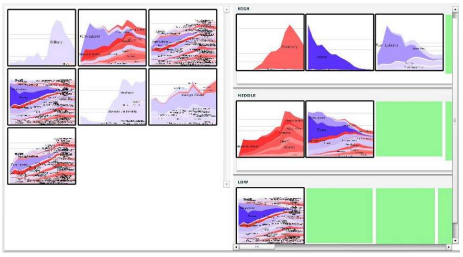

Figure 1: *Rating Interface.*

3. Results and Discussion

Table 1 shows the evaluation results of implicit vs explicit interest indicators. It includes the average and, enclosed in parentheses, the standard deviation.

Table 2 shows the average of ranks according to each rating group for all views produced for questions 1 and 2. The Kruskal Wallis technique calculates the mean of ranks of the data rather than numeric values. The ranks are found by ordering the data from smallest to largest across all groups, and taking the numeric index of this ordering. The number of data here is the total number of produced views for both questions, i.e., 281 views. Table 2 shows the rank average belonging to each group. The rank average of dwell time in high, middle, and low groups are in ascending order while those for revisits and mouse activity are not in ascending order.

We used the median and distribution of each indicator using a Kruskal-Wallis test (based on the 99% confidence interval). Our results are as follows:

Table 1: *Evaluation results of the implicit VS explicit interest indicators.*

question	usefulness	dwell time	revisits	mouse
1	HIGH	24.20 (22.29)	1.86 (1.92)	754.64 (670.21)
	MIDDLE	12.71 (12.73)	1.39 (2.74)	420.39 (412.54)
	LOW	9.16 (10.48)	0.60 (0.99)	246.25 (300.67)
2	HIGH	23.75 (17.08)	1.19 (1.74)	589.20 (415.83)
	MIDDLE	14.91 (17.32)	0.39 (0.92)	468.85 (631.86)
	LOW	11.43 (12.74)	0.30 (0.72)	277.08 (371.51)

Table 2: *Rank average according to each group for all views produced for questions 1 and 2*

	HIGH	MIDDLE	LOW
Dwell Time	176.1	115.5	94.1
Revisits	164.9	119.9	112.3
Mouse Activity	173.6	125.7	90.2

(dwell time|revisits|mouse) vs explicit rating: the null hypothesis was rejected, meaning that the median values for each explicit rating group differed.

Three indicators vs (questions|users): the null hypothesis of each indicator was accepted for: 1. both questions, and 2. all 14 users, indicating that our findings are universal, i.e, independent to the questions and users.

Our conclusion is that *dwell time*, *revisits*, and *mouse activity* on a view are good interest indicators. Our future work includes searching for a prediction function that accurately predicts explicit interest.

References

[CLWB01] CLAYPOOL M., LE P., WASED M., BROWN D.: Implicit interest indicators. In *IUI '01: Proceedings of the 6th international conference on Intelligent user interfaces* (New York, NY, USA, 2001), ACM, pp. 33–40. 1

[HVW09] HEER J., VIÉGAS F. B., WATTENBERG M.: Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM 52*, 1 (2009), 87–97. 1

[RS01] RAFTER R., SMYTH B.: Passive profiling from server logs in an online recruitment environment. In *IJCAI Workshop on Intelligent Techniques for Web Personalization (ITWP 2001)* (USA, 2001), pp. 35–41. 1

[WK06] WATTENBERG M., KRISS J.: Designing for social data analysis. *Visualization and Computer Graphics, IEEE Transactions on 12*, 4 (july-aug. 2006), 549 –557. 1

[ZF98] ZHOU M. X., FEINER S. K.: Visual task characterization for automated visual discourse synthesis. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1998), ACM Press/Addison-Wesley Publishing Co., pp. 392–399. 1