IEEE Visualization Conference, Poster Extended Abstract. Phoenix, AZ, October 2017

Applying Multi-Player Rating Schemes to Manage User Studies of Visual Analytics Designs

Salman Mahmood, Klaus Mueller*
Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University

Abstract

Elo is a popular rating scheme in chess and online games. We propose a modified version of the Elo rating system to compare visualization designs in an efficient and asynchronous manner. Our method allows researchers to evolve the design of a visualization layout and monitor its performance. We study our method for testing small multiple designs. Small multiples is a very popular technique to reduce clutter in visualizations. However, they can be difficult to interpret for people who are not entirely familiar with them. We designed various exploded view methods to make small multiples comprehensible and then compared them using our adaptation of the Elo rating system.

1 Introduction

When evaluating designs for a data visualization researchers have to choose between many design choices. Visualization designs evolve over time and designers are interested in seeing how these updated versions compete with older versions. Our method is an asynchronous framework that allows designers to inject new versions into the solution space and visually monitor their success.

Testing designs is difficult because the human brain is the only appropriate platform for testing visualization techniques. It can be difficult to recruit enough participants to run a proper user study for visual analytics design. The widespread connectivity of people has opened the door for recruiting humans for user studies. Crowdsourcing is an effective technique which uses monetary incentives to recruit people. In this way, a near instantaneous human-based assessment of competing visualizations can be achieved. We propose a modified version of the Elo ranking system to monitor the competitiveness of the competing evolving design in real time. To demonstrate the use of our progressive rating scheme we use an evolving design of exploded views for scatterplot visualization. We created three designs for multi-clustered scatterplots. The designs are then tweaked to see if the changes improve the designs or not. The exploded view designs were created as part of an ongoing project that is subject to a future publication.

2 Elo Rating Scheme

Elo is a very popular rating scheme. It was introduced as a chess rating system by Aprad Elo [2]. Variants of the rating scheme are still used to rate chess players. After each match, the rating of a player goes up or down depending on the result and the ratings of the competitors. For a competition between two players, we compute the probability of winning for each player. For players i and j the expected probability for i to win is defined as follows:

$$E_i = \frac{1}{1+10^{\frac{(R_i+R_j)}{400}}} \tag{1}$$

*e-mail: {samahmood, mueller}@cs.stonybrook.edu

where R_i and R_j are the ratings of the players i and j. The final rating of the player is calculated using the following formula:

$$R_i = R_i + K(S_i - E_i) \tag{2}$$

here K is an attenuation factor that determines the weight that should be given to each players performance and S_i represents the actual result of the competition. Therefore, if the expected probability E_i is greater than the result S_i (i.e the player is expected to do better than the actual outcome), the player's ratings will drop and vice versa. The magnitude of the change is dependent on two factors. Firstly the difference between the expected result and the actual result and second the value of K. A large value of K will make the ratings more sensitive to wins and losses. In chess competitions, the value of K is kept high for players with lower ratings and low for players with higher ratings. In our method K represents the reliability of the user, More reliable users will have a larger value of K.

3 Example Use: Exploded view

Small multiples is a displacement technique that has been widely used in data analysis. However, they can be difficult to understand for people who are not familiar with them. An exploded view is a diagram in which the different components of an object are suspended in the space around the object. Giving the impression that the object is midway through an explosion. Exploded view diagrams are very useful in capturing the relationships between different objects as well as revealing the assembly and structure of an object. We feel that an exploded view is a very natural solution to make the small multiples easier to understand. We use a force directed layout scheme to determine the positions of the components during the explosion [1].

We designed 3 different methods for the exploded view of small multiples (see Fig. 1 for illustrations). They are as follows:

Trajectory, in this method the different components of the graph fall into their final positions following a linear trajectory, calculated using the force configuration algorithm.

Firework, this method has three phases the first phase is an implosion, the points gather at the center to form a small ball, in the second phase the small ball then moves to the final position and in the third phase the components explode like a firework.

Cluster, we cluster the different components of the data using the hierarchical clustering method. We find the number of clusters by using the elbow method. The Cluster Explosion is divided into two phases, in the first phase of the explosion, the components that belong to the same cluster move together. In the second part, we explode the clusters and separate the data points individually.

More detail on our exploded views paradigm for visualization will be presented in a future publication

4 Elo-Based User Survey Monitoring

In the user survey, we will compare the different designs listed in Sect. 3 and find the best design using the modified Elo rating scheme. We used the Amazon Mechanical Turk (AMT) to perform our survey. The task was listed as a Survey Link task, where the participant is provided with a link to a website. After completion of the survey, the participant is given a code which they enter into the AMT.

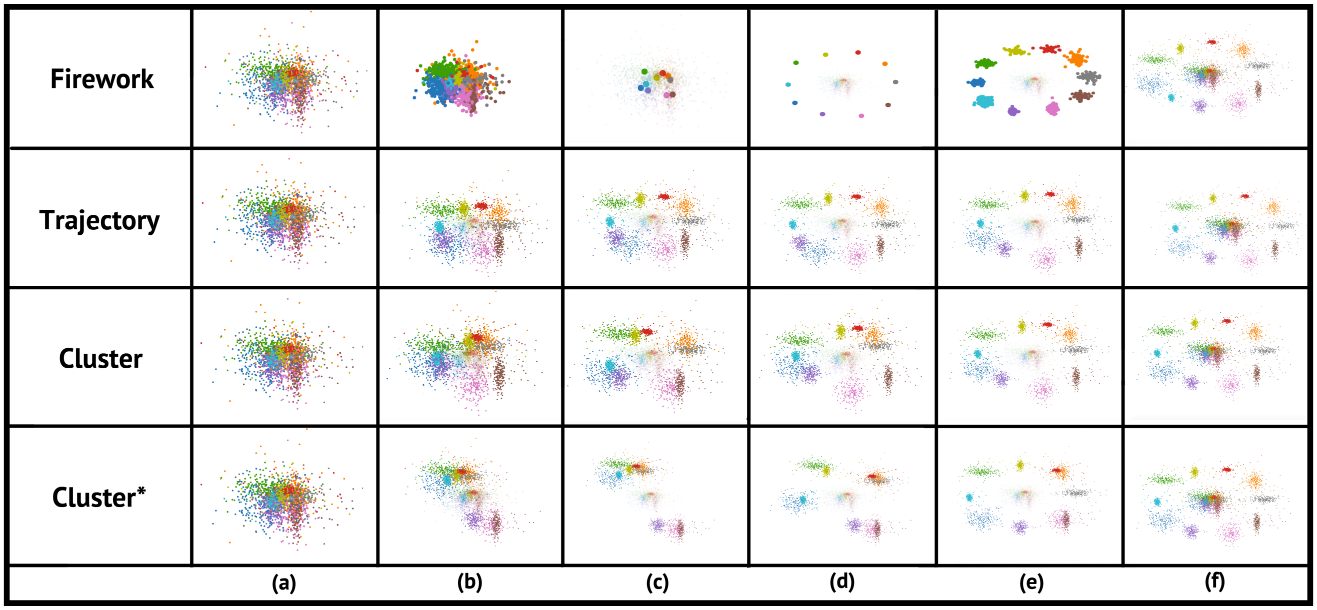

Figure 1: Shows the different phases of the explosion The actual points are still kept in the center. The red line indicates the point at which the cluster method (row 3) was enhanced to cluster*(row 4)

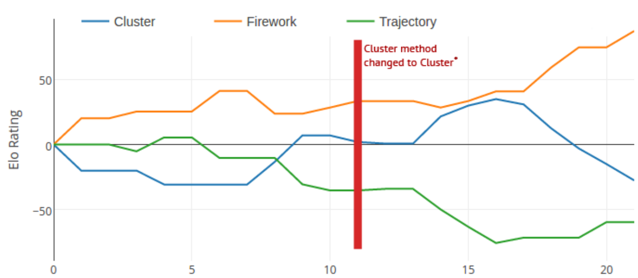

Figure 2: Changes in the ratings of the exploded view designs. It can be clearly seen that the effectiveness of the Cluster method has decreased after the change was introduced (red bar). After a brief short-term increase, its long-term competitiveness falls drastically while the Fireworks method dominates. Conversely, The Trajectory method never really catches on.

4.1 Experiment Setup

In the survey, each participant was asked to compare three pairs of Exploded View designs. In a comparison, they were asked to solve multiple choice questions (MCQ) by using the exploded view. The questions were related to the outliers, clusters and density of the data. The participants were asked to answers questions for each comparison until they got two questions right for each comparison. When evaluating a match between two exploded view methods we took into consideration the accuracy and the preference of the participant. The accuracy was determined by the results of the MCQs. After they had viewed the two designs, they were asked which design they would prefer, with the option of selecting either one of the designs or both. They were shown animated GIF files to facilitate recollection of the designs. The accuracy and preference were then combined to find the winner of each comparison.

We also use the number of questions answered to estimate the reliability of the participant. Each participant had to answer a total of 6 questions correctly so a perfectly reliable participant would answer 6 questions and a participant that answers questions randomly would answer 24 questions approximately. We use the following function to determine the reliability of the participant:

$$K = 32 * max(0, 1 - \frac{n-6}{18}) \quad (3)$$

where n is the number of questions answered by the participant.

4.2 Experiment and Results

The experiment was divided into two parts. In the first part we recruited 11 participants. The responses of the different participants are interleaved in this example. The results of the survey are shown in Fig. 2. It can be seen that up to the midway point (represented by the red line) the firework method is the best followed by the cluster method, whereas the trajectory method does not do too well. We made a small change to the Cluster method (row 3 in Fig. 1) to see if that improves our results. The explosion is divided into multiple phases. In every phase we double the number of clusters (see row 4 in Fig. 1). To test the performance of the Cluster method after the changes we recruited another 10 participants to measure the change and found that the changes reduced the effectiveness of the method (see the eventual drop of its rating curve in Fig. 2). This improved the ranking of the Fireworks method even more.

5 Conclusion

This paper presents a pairwise comparison technique for visual analytic design with many alternatives. We use the Elo ranking system to rate multiple designs created for small multiples implementations. Elo can rank different designs using a small number of participants. We show its potential to monitor evolving designs.

6 Acknowledgment

This research was partially supported by NSF grant IIS 1527200

References

[1] S. Bruckner and M. E. Groller. Exploded views for volume data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006.

[2] A. E. Elo. *The rating of chessplayers, past and present*. Arco Pub., 1978.