Analysts aren't machines: inferring frustration through visualization interaction

Lane Harrison *　Wenwen Dou　Aidong Lu　William Ribarsky　Xiaoyu Wang

Computer Science
UNC-Charlotte

Abstract

Recent work in visual analytics has explored the extent to which information regarding analyst action and reasoning can be inferred from interaction. However, these methods typically rely on humans instead of automatic extraction techniques. Futhermore, there is little discussion regarding the role of user frustration when interacting with a visual interface. We demonstrate that automatic extraction of user frustration is possible given action-level visualization interaction logs. An experiment is described which collects data that accurately reflects user emotion transitions and corresponding interaction sequences. This data is then used in building Hidden Markov Models (HMMs) which statistically connect interaction events with frustration. The capabilities of HMMs in predicting user frustration are tested using standard machine learning evaluation methods. The resulting classifer serves as a suitable predictor of user frustration that performs similarly across different users and datasets.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 Introduction

We hypothesize that human-to-visualization interaction patterns can be used to automatically infer and predict user frustration. In human-to-human discourse, people are often able to detect frustration in others by observing the way that others interact with them. Similarly, in human-to-visualization discourse, we wish to see if a system can make similar inferences. Complicating the problem is the fact that the system can only detect a limited number of the user's actions, mainly through the keyboard and mouse. While other input devices can be used, it is not practical to assume that such devices are available in practice. Our work demonstrates that through a limited number of interactions, a statistical model can be built that predicts frustration.

To accomplish this, we collect ground-truth data that includes visualization interaction logs and emotional state transition timelines. We then train Hidden Markov Models using this data and evaluate their predictive capabilities using evaluation methods from the machine learning community.

Research in psychology indicates that emotions sometimes serve as a strong influence in even the most critical analytical situations [6]. Knowledge of emotions such as frustration have also been used to help disambiguate human intent in user-interface studies [5]. Additionally, aspects of human computer interaction that could affect analysis (such as intent and frustration) have been identified as important areas for research in visual analytics [7]. Because of the impact of emotions in human decision making, the ability to detect emotions such as frustration could prove useful to visual analytics systems.

*e-mail: ltharri1@uncc.edu

IEEE Symposium on Visual Analytics Science and Technology
October 23 - 28, Providence, Rhode Island, USA
978-1-4673-0013-1/11/$26.00 ©2011 IEEE

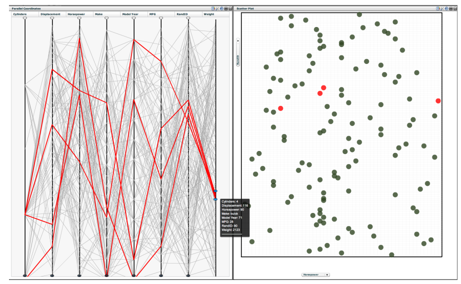

Figure 1: An overview of the linked-view visualization system participants used

Previous work in analytics has explored the extent to which high-level actions can be inferred from analyst interaction [6], [2]. Other work investigated the abilities of novice users to infer the reasoning strategies of expert analysts [3], and to assist expert analysts in recalling their own reasoning processes. Similarly, our work seeks to explore the extent to which interaction logs can be used to infer the emotional state of analysts. In contrast to previous work, we propose a method of automatically predicting engagement and frustration without the aid of human coders.

2 Experiment

To build a predictive model of frustration, we first gather ground-truth data that accurately reflects the emotional state transitions and interaction sequences of humans using a visualization system shown in figure 1.

A scatterplot view was included since many users were likely to be familiar with them and their interactions are easily taught and understood. A parallel coordinates view was also included to both increase the number of interaction events to log and to serve as a particularly useful view for some of the tasks (for example, seeing trends across dimensions). The views were coordinated in that selections in one view are also selected in the other view. Coordination helped users see the relationship between scatterplots and parallel coordinates.

Participants were asked to complete three to four tasks for both the cereal and cars datasets that are widely used in visualization research. While the participants are completing the tasks, several types of interactions are logged. These include actions like selection, changing axes, and resetting the visualizations. A complete list is given in table 1.

Video recording is used to capture the participant's face and screen to assist in building an emotion transition timeline after all tasks are completed. Specifically, after all tasks are completed, the user views the videos (of the face and screen) with an investigator and indicates points at which they become frustrated or engaged (transition out of frustration). It should be noted that although it has been shown that users in specific situations can fail to accurately express what they feel via self-reporting, self-reporting remains accepted as standard practice within the affective computing commu-

Table 1: The ten interaction events logged by our system (PC: Parallel Coordinates; SC: Scatterplot

PC-Tooltip	Mouse-over to display detailed information in parallel coordinates
SC-Tooltip	Mouse-over to display detailed information in scatterplot
SC-Click	Highlight an item in the scatterplot
PC-Click	Highlight an item in parallel coordinates
Change SC Axes	Select new scatterplot axes
Change PC Axes	Rearrange axes in parallel coordinates
PC minbar moved	Move the bottom slider in a parallel coordinates axis
PC maxbar moved	Move the top slider in a parallel coordinates axis
Opening WorkSpace	Open a new display
Closing WorkSpace	Close a display

nity [4]. Using these emotion-transition timelines and interaction logs, we are able to build a predictive model of user frustration.

2.1 Hypotheses

For our experiment, we identified the following hypotheses to test:

- Frustration influences a user's interaction patterns with a visualization system.
- Given training data, a user's affective state can be automatically predicted from their interaction patterns.
- Using the same visualizations and same tasks, interaction patterns that reflect emotional states are generalizable across users.
- Using the same visualizations and similar tasks, interaction patterns that reflect emotional states are generalizable across datasets.

2.2 Procedure

The primary purpose of this experiment was to collect ground-truth data that can be used to build a predictive model. As such, the experiment can be divided into two portions: that of task-completion to collect interaction data and that of video mark-up to generate emotion transition timelines.

After participants completed consent and demographics forms, they were given an overview of the visualization system (see Figure 1) during which all of the possible interactions were explained. Additionally, participants were reminded about concepts such as correlations and outliers to better prepare them for the tasks.

After the overview, the participants were given the task sheets to look over and instructed to ask any questions they might have. Tasks included questions that increase in difficulty. For example, a simple task was "In what year was the car with the highest MPG made?". While a more difficult task was "What is the lightest car with the highest horsepower?". Next, the participants were given the Self-Assessment Manikin (SAM) test, which is used in psychology research to determine a person's emotional state [1]. The SAM was also given as a post test.

To facilitate the participants' focus on the questions, they were not asked to report their emotional state during the tasks. When the participant finished the assigned tasks, the investigator displayed the recordings of both the participant's face and of the screen. Then the investigator and the participant reviewed the videos simultaneously and identified points at which the participant transitioned from engaged or interested to frustrated or confused and vice versa.

Table 2: Results of leave-one-out cross validation across users

Metric	Cars	Cereal
Accuracy	67.71%	67.56%
Error	32.29%	32.44%
Sensitivity	61.28%	56.89%
Specificity	75.23%	80.44%
Avg true-positive	189.1	161.8
Avg true-negative	198.9	189.6
Avg false-positive	65.5	46.1
Avg false-negative	119.5	122.6

Table 3: Results of cross validation across datasets

Metric	Cars	Cereal
Accuracy	67.28%	67.47%
Error	32.72%	32.53%
Sensitivity	60.82%	56.79%
Specificity	74.81%	80.36%
Avg true-positive	187.7	161.5
Avg true-negative	197.8	189.4
Avg false-positive	66.6	46.3
Avg false-negative	120.9	122.9

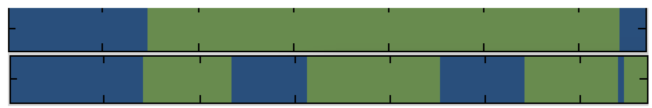

Figure 2: (Top) the predicted state transitions for a user and (bottom) the corresponding actual state transitions (green is frustrated, blue is engaged). The accuracy of these predictions was 70%

3 Results

The user study provided us with interaction logs for which each interaction event has a corresponding affective state. Using part of this data for training and part for testing (cross validation), we trained Hidden Markov Models and tested their predictive capabilities on data in which we remove the participants' reported affective states. This was repeated to test our hypotheses. Tables 2 and 3 refer to the hypothesis regarding generalizability across users and data, respectively. The results indicate that the interactions logged can be used to infer frustration or engagement across datasets with nearly 70% accuracy. Frustration was more accurately inferred than engagement, with engagement being classified as low as 57% and frustration as high as 80%. Refer to table 2 for additional results.

References

[1] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59, 1994.

[2] S. Chaiken. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 1980.

[3] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *Computer Graphics and Applications, IEEE*, 2009.

[4] R. W. Picard and S. B. Daily. Evaluating affective interactions: Alternatives to asking what users feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*. ACM, 2005.

[5] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1175–1191.

[6] J. Richard Heuer. *The Psychology of Intellligence Analysis*. 1999.

[7] J. S. Yi, Y. ah Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 2007.