Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Constructing Models of User and Task Characteristics from Eye Gaze Data for User-Adaptive Information Highlighting

Matthew Gingerich and Cristina Conati
University of British Columbia
{majugi, conati}@cs.ubc.ca

Abstract

A user-adaptive information visualization system capable of learning models of users and the visualization tasks they perform could provide interventions optimized for helping specific users in specific task contexts. In this paper, we investigate the accuracy of predicting visualization tasks, user performance on tasks, and user traits from gaze data. We show that predictions made with a logistic regression model are significantly better than a baseline classifier, with particularly strong results for predicting task type and user performance. Furthermore, we compare classifiers built with interface-independent and interface-dependent features, and show that the interface-independent features are comparable or superior to interface-dependent ones. Finally, we discuss how the accuracy of predictive models is affected if they are trained with data from trials that had highlighting interventions added to the visualization.

Introduction

Research in information visualization has always had a strong focus on understanding how cognitive and perceptual processes interact with visualizations. In recent years, there has been increased attention in investigating the impact of users' *individual differences* on interactions with visualizations. Researchers have found that several cognitive abilities and personality traits can impact the effectiveness of visualizations (e.g. Ziemkiewicz et al., 2011, Toker et al., 2013). These findings suggest that personalized visualizations may offer benefits to users by adapting to their individual traits.

The goal of our current work is to explore what value gaze data can have for an adaptive visualization that actively learns the traits of its users and presents customized highlighting of information to meet their individual needs. Gaze patterns are of interest both because they are tightly linked to information processing (Just and Carpenter, 1976), and because gaze data can be obtained for non-interactive visualizations. Earlier work showed that three different cognitive abilities (perceptual speed, visual working memory and verbal working memory), task performance, and task type can be estimated (albeit with varying degrees of accuracy) from gaze patterns during visualization tasks with bar graph and radar graph visualizations (Steichen et al., 2014a). Our work complements these previous findings as follows.

First, we verify whether we can obtain similar results for predicting the same user and task properties while users interact with bar graphs visualizing more complex datasets. The inclusion of task performance as a target for prediction is intended to address the question of when adaptive interventions should be displayed: users with low predicted performance stand to benefit more from additional support. The prediction of user cognitive abilities and task type can inform the decision of which interventions should be displayed, since prior work has correlated these characteristics with effects on specific aspects of visualization processing (Carenini et al., 2014).

Second, we add the prediction of a user's locus of control, a personality trait that has been shown to impact visualization performance (e.g. Ziemkiewicz et al., 2011).

Third, we compare classifiers built with interface-independent and interface-dependent features, in order to assess the extent to which these predictive models require having detailed information on the presented visualization.

Finally, we investigate how model accuracy is affected if models are trained with data from tasks that had different types of *highlighting interventions* added to the visualization. These interventions were designed to highlight graph bars that were relevant to perform the given task. Interventions could eventually be used to provide adaptive support by dynamically redirecting the user's attention to different subsets of the visualized data as needed (e.g. when the visualization is used together with a verbal description that discusses different aspects of a dataset, Carenini et al., 2014).

The rest of this paper presents related work, followed by a summary of the user study that generated our datasets. Next, we illustrate the classification experiments that were run to predict the aforementioned user and task properties. Finally, we discuss our results and conclusions.

Related Work

Our work on user-adaptive visualizations draws from research in three related areas: analyzing influences of user traits on visualization effectiveness, user modeling, and the use of eye-tracking to build user and task models.

The influence of user traits on the effectiveness of visualizations has been studied for both cognitive abilities and personality-based traits. Perceptual speed, visual working memory, and verbal working memory cognitive abilities were found to influence both performance with and preferences for visualizations (Conati and Maclaren, 2008), (Velez et al., 2005) (Toker et al., 2012). The locus of control personality trait was found to influence performance on visualization tasks (Ziemkiewicz et al., 2011).

Studies linking user traits to visualization effectiveness motivate the need to estimate those traits during visualization use. Several researchers have approached this task by tracking user interface actions. For instance, Grawemeyer (2006) and Mouine and Lapalme (2012) recorded user selections among alternative visualizations to recommend visualizations in subsequent tasks. Gotz and Wen (2009) track suboptimal user interaction patterns to recommend alternative visualizations for the current task. (Ahn and Brusilovsky, 2013) track a history of user search terms to customize the display of exploratory search results.

Gaze data has been shown to be a valuable source of information for user modeling in various domains. Eivazi and Bednarik (2011) used gaze data to predict user strategies when solving a puzzle game. Kardan and Conati (2013) and Bondareva et al. (2013) use gaze to predict student learning with educational software, while Jaques et al. (2014) leverages it for affect prediction. Liu et al. (2009) predict skill level differences between users in collaborative tasks.

In information visualization, user modeling with gaze data was explored by Steichen et al. (2014a), and Toker et al. (2014). Steichen et al. found that task type and complexity, user performance, and three user cognitive abilities (perceptual speed, visual working memory and verbal working memory) could be classified with accuracies significantly above a majority class baseline. Their work used simple bar graph visualizations with at most three data points per series. In this paper, we use data from a study that involved more complex bar graphs (doubling the maximum data points per series) and added highlighting interventions to the graphs. We also add the classification of a user's locus of control, and test classification accuracy with gaze features built upon interface-independent areas of interest (AOIs). Toker et al. (2014) used the same dataset leveraged in this paper to model users' skill acquisition. They also looked at the performance of interface-independent AOIs and found that they did not perform as well as AOIs based on specific interface features for predicting skill acquisition. In this paper, we extend the work of Toker et al. on interface-independent AOIs to the classification of task type, user performance, and user cognitive traits.

User Study

The gaze data used in this paper was collected during a user study that investigated both the effectiveness of four highlighting interventions, as well as how this effectiveness is impacted by task complexity and different user traits (Carenini et al., 2014). In the study, 62 participants between the ages of 18 to 42 were given bar graph visualizations along with textual questions on the displayed data (see Figure 1). There were 39 female participants and 23 males. The experimental software was fully automated and ran in a web-browser. User gaze was captured with a Tobii T120 eye-tracker, embedded in the computer monitor.

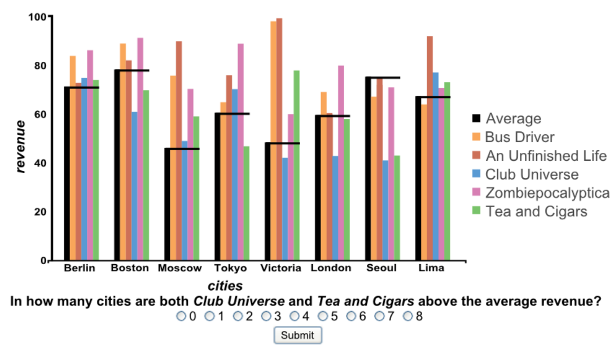

Figure 1. *The study interface showing a task without highlighting interventions (components have been scaled for printing).*

Task complexity was varied by having subjects perform 2 different types of tasks, chosen from a standard set of primitive data analysis tasks (Amar et al., 2005). The first task type was Retrieve Value (RV), one of the simplest task types in (Amar et al., 2005), which in the study consisted of retrieving the value for a specific individual in the dataset and comparing it against the group average (e.g., "Is John's grade in *Philosophy* above the class average?").

The second, more complex task type, was Compute Derived Value (CDV). The CDV task in the study required users to first perform a set of comparisons, and then compute an aggregate of the comparison outcomes (e.g., Figure 1). All tasks involved six data points and eight series elements.

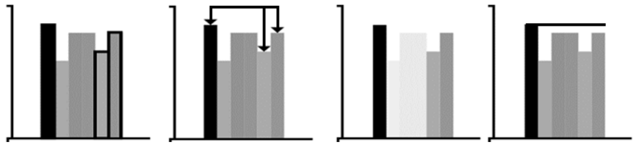

Figure 2. *Highlighting interventions. Left to right: bolding, connected arrows, de-emphasis, and reference line.*

Each intervention evaluated in the study (shown in Figure 2) was designed to highlight graph bars that were relevant to answer the current question, to guide a user's focus to a specific subset of the visualized data while still retaining the overall context of the data as a whole (Few, 2009). The *Bolding* intervention draws a thickened box around the relevant bars; *De-Emphasis* fades all non-relevant bars; *Average Reference Lines* draws a horizontal line from the top of the left-most bar (representing the average) to the last relevant bar; *Connected Arrows* involves a series of connected arrows pointing downwards to the relevant bars. Participants began by completing a set of tests that measured the 5 user characteristics evaluated in the study, selected because they had been previously shown to influence user performance or satisfaction with visualizations. They included: (1) *Perceptual speed*, a measure of speed when performing simple perceptual tasks (Ekstrom, 1996); (2) *Visual Working Memory*, a measure of storage and manipulation capacity of visual and spatial information (Fukuda and Vogel, 2009); (3) *Verbal Working Memory*, a measure of storage and manipulation capacity of verbal information (Turner and Engle, 1989); (4) *Bar Graph Expertise*, a self-reported measure of a user's experience with using bar graphs; and (5) *Locus of Control*, a personality trait measuring whether individuals tend to take responsibility for their circumstances or blame them on external factors. Next, each participant performed each of the two task types (RV & CDV) with each of the 4 interventions as well as *No Intervention* as a baseline for comparison, in a fully randomized manner. Additional details about the user study protocol and references to all the standardized tests used to assess user characteristics can be found in (Toker et al., 2013).

Design of Classification Experiments

Classification Targets

We used the dataset collected from the study described in the previous section to build classifiers for the following classification targets: task type, user performance, visual working memory (visual WM), verbal working memory (verbal WM), locus of control, perceptual speed (PS), and expertise with bar graphs. All classifiers leverage only gaze features as inputs and predict binary labels. Binary labels were created from median splits of continuous traits.

The task type could be either RV or CDV representing, respectively, tasks of lower or higher complexity. User performance represents the time required by users to complete tasks, where longer completion times are indicative of lower performance.

Eye Tracking Features

Eye tracking data consists of fixations (i.e., gaze points on the screen), and saccades (i.e., quick movements between fixations). For our classification experiments, we process the raw gaze data into a set of basic measures (fixation number, rate and duration, saccade length, and absolute/relative saccade angles) calculated using the open-source Eye Movement Data Analysis Toolkit (EMDAT). EMDAT converted these basic measures into individual eye-tracking features by calculating summary statistics such as their mean, sum, and standard deviation. Additional features were obtained by defining specific areas of interest (AOIs) within the interface and then calculating statistics on basic gaze measures restricted to these AOI. Transitions between AOIs are also captured by AOI-specific features.

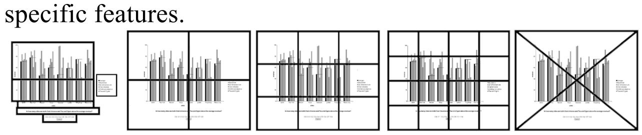

Figure 3. *Areas of interest (AOIs). Left to right: Custom AOIs, 2×2 Grid, 3×3 Grid, 4×4 Grid, and X Grid.*

To test how much knowledge of the current interface is necessary for accurate classification of our targets, we defined different types of AOI. The Custom AOI set includes six interface-specific AOIs manually defined to cover the functional elements of the interface used in the study (e.g. bar graph labels, legend, top part of the bars in the graph, question text, input area). Following (Toker et al. 2014), four additional, interface-independent, sets of AOIs were created based on the geometric grids shown in Figure 3 (Generic AOIs from now on). AOI-dependent features were supplemental to the AOI-independent summative gaze features. All classification experiments were repeated with each of the five AOI sets and with a set including only AOI-independent features. The set of only AOI-independent features contained 14 features, while the other feature sets ranged in size from 78 features to 654 features.

Classifier and Evaluation Details

To simulate the temporal evolution of the data that would be available to an adaptive system, we created "partial" datasets by generating features from percentages of the interaction time for each trial (e.g. the first 10% of a trial's data, 20%, etc.). For each classification target and "partial" dataset, separate classification experiments were run for each possible combination of a displayed intervention and an AOI set used for feature generation.

All classification experiments tested the accuracy of using logistic regression and correlation-based feature selection (Hall, 1999) for learning models. Logistic regression was selected after testing it alongside random forest, support vector machine, multilayer perceptron, decision tree, and naïve Bayes classifiers because it had the highest overall accuracy. Models were trained using the WEKA machine learning toolkit (Hall et al., 2009), and evaluated via ten runs of 10-fold cross-validation by comparison against a majority class baseline classifier.

In order to summarize the multitude of results from these classification tasks, we run three-way, 2 (classifier type) by 6 (AOI type) by 5 (intervention type), repeated measures ANOVAs for each classification target with classification accuracy as the independent measure and use these to (i) identify significant differences between the performance of logistic regression and the baseline; (ii) systematically compare the effect of interventions and AOIs on classification results. Bonferroni adjustments are applied to correct for both the multiple statistical comparisons conducted due to the eight separate ANOVAs, as well as pairwise tests used to interpret effects within each ANOVA. Reported effects are all significant at the $p<0.05$ level after adjustments.

Results

Classifier Effects

For all our classification targets except visualization expertise and locus of control, main effects of classifier in the ANOVAs indicate that logistic regression performs significantly better than the baseline. Because of the poor performance for visualization expertise (which mirror the findings from Steichen et al., 2013) and locus of control, these targets are dropped from the rest of the paper. As Table 1 shows, the mean accuracy of the logistic classifier is above 90% for task type and 85% for completion time.

Table 1. *Main effects of classifier type on classification.*

Classification Target	Logistic Mean (%) [Std. Dev.]	Baseline Mean (%) [Std. Dev.]
Task Type	91.5 [6.38]	51.6 [2.00]
Completion Time	88.2 [3.50]	73.6 [6.55]
Perceptual Speed	59.9 [3.90]	51.2 [0.92]
Visual WM	56.0 [3.52]	54.4 [0.84]
Verbal WM	61.6 [2.55]	59.8 [0.81]

Accuracies for user characteristics are lower, consistent with the findings reported by Steichen et al. (2013). It should be noted, however, that the means reported in Table 1 are obtained by averaging over the results from all the tested AOI feature sets, interventions, and varying amounts of observed data. Thus, these means are conservative estimates of the accuracies that could be achieved in practice since less informative feature sets would not be used.

Effects of AOI Type and Interventions

Our ANOVA results show significant main effects for both AOI type and interventions on classification accuracy for all classification targets. These effects can be interpreted using pairwise comparisons between levels of each factor. Table 2 and Table 3 summarize all pairwise comparisons between, respectively, AOI types and interventions. In these tables, underlines group levels for which there are no statistically significant differences in accuracy. For example, Table 2 shows that Custom AOIs and X Grid AOIs do not have significantly different accuracy in the classification of perceptual speed.

Table 2. *Effects of AOI type on classification. See text for details.*

Classification Target	AOI Type (Lowest Classification Accuracy to Highest)
Perceptual Speed	No AOIs < Custom < X < 4×4 < 3×3 < 2×2
Visual WM	No AOIs < 3×3 < Custom < X < 2×2 < 4×4
Verbal WM	No AOIs < 4×4 < 3×3 < X < 2×2 < Custom
Task Type	No AOIs < X < Custom < 3×3 < 4×4 < 2×2
Completion Time	No AOIs < Custom < X < 4×4 < 2×2 < 3×3

Main Effects of AOI Type

Several trends are visible from the pairwise comparisons in Table 3. The No AOI feature set consistently performs worse than the AOI-based sets, thus showing the value of the finer-grained information provided by the AOI-based measure. On the other hand, the Custom AOIs are generally not better than generic feature sets. In fact, for each classification target except verbal working memory, there is a generic AOI feature set with significantly higher accuracy than the Custom AOI. These findings suggest that manually isolating functional regions in the interface may not always be necessary for creating informative AOIs, and thus provide encouraging, although preliminary, evidence that gaze-based classifiers can be built without *a priori* information on the target visualizations. This result is considered preliminary as its generality has yet to be tested on a wide variety of visualizations or data distributions.

Verbal WM and visual WM share several trends with respect to their accuracy with each type of AOI (e.g. the 2×2 grids were better than both the 3×3 grids and the X grids, which were in turn better than having no AOIs). The similarity in these trends could be a reflection of the similarity of the role of these two traits in task processing.

Aside from the trends noted above, the relative accuracies obtained with each set of AOIs vary among classification targets. This finding can be attributed to each target having distinct influences on the way users interact with the interface, which in turn influences which interface elements and related gaze patterns are most predictive. For example, task type influences the number and position of bars that participants must examine, while verbal WM influences how long users can retain information given in the question text, the graph labels, and the legend. The varied performance of different AOI across targets may reflect their different abilities in capturing the relevant interface elements and related attention patterns. For example, in the 4×4 Grid AOIs, three of the grid cells lie near the top of bars in the bar graph and may collectively capture the majority of visual comparisons of bar length. Features from these three AOIs were often selected in the classification of visual working memory.

Main Effects of Interventions

As was the case for feature sets, the main effects of interventions, summarized in Table 3, vary with each classification experiment. Classification accuracy with the *None* condition (in which no highlighting is presented, further discussed in the next section), is of particular interest for practical applications because it emulates the scenario in which an adaptive system must decide if and how interventions should be displayed. In contrast, the prediction accuracies in tasks with interventions provide data for how well a system could continue to refine its user and task models after presenting interventions.

Table 3. *Effects of interventions on classification. The average reference line intervention is abbreviated as 'Line' and the connected arrow intervention is abbreviated as 'Arrow'.*

Classification Target	**Intervention (Lowest Classification Accuracy to Highest)**
Perceptual Speed	Bold < None < De-emphasis < Arrow < Line
Visual WM	De-emphasis < Line < Arrow < None < Bold
Verbal WM	De-emphasis < None < Arrow < Line < Bold
Task Type	None < Line < Bold < Arrow < De-emphasis
Completion Time	None < Line < Bold < Arrow < De-emphasis

The effects in Table 3 show that classification accuracy with the *None* condition is often worse than accuracy with an intervention. For every classification target there is at least one intervention that correlates with statistically significant improvements in predictions. For instance, for visual working memory classification, significantly better predictions can be obtained with the Bolding intervention (59% mean accuracy) than with no intervention (57% mean accuracy), but presenting other interventions reduced the accuracy of predictions. This variation in prediction accuracies across interventions may be due to the fact that, in some cases, interventions may make classification more difficult by reducing the differences in user gaze behaviors between the two groups to be predicted (e.g. helpful interventions may make the gaze behavior of low perceptual speed users closer to that of their high perceptual speed counterparts).

In addition to the main effects reported above, there were interaction effects between the AOI and intervention factors, as well as classifier type. One general implication of these results is that the effect of using a partic-ular feature set is dependent on the intervention displayed during trials, but the specific interactions are difficult to interpret owing to the large number of conditions that could be individually considered. In the interest of brevity, in the next section we focus on classification results for the *No Intervention* condition, which is the condition of most direct interest for providing adaptive help and the condition that is most directly comparable to the work of Steichen et al. (2013).

Best Obtained Models with No Intervention

Looking only at tasks in which users received no interventions, Table 4 reports, for each classification target, the mean accuracy across all datasets achieved by the best performing AOI set. To give a better sense of how classification accuracy changes with the amount of data available, Figure 4 shows the trends in accuracy over time for some of the classification targets. Recall that the time factor is simulated by creating "partial observation" datasets that sample data from the start of task trials.

Table 4. *Best obtainable classifiers with no intervention.*

Classification Target	**Baseline Accuracy**		**Logistic Regression Accuracy**		
	Mean	**Std. Dev.**	**Best AOIs**	**Mean**	**Std. Dev.**
Task Type	51.0	0.40	3 × 3	92.3	6.28
Compl. Time	65.1	0.77	3 × 3	86.2	2.45
Perceptual Sp.	50.9	0.19	2 × 2	61.4	3.32
Visual WM	54.3	0.43	4 × 4	61.5	1.91
Verbal WM	59.5	0.43	4 × 4	61.7	1.67

Classification accuracy for task type increases as more data becomes available throughout the trial, but it already reaches 81.7% after observing only 10% of the data. This result has direct implications for providing adaptive support to users. For instance, Carenini et al. (2014) found that certain highlighting interventions had better performance on RV than CDV tasks, thus early prediction of user task could help an adaptive visualization select the most appropriate intervention as soon as it is needed.

User performance classification, measured by completion time, complements task type classification by identifying occasions in which users are most in need of support. Accuracy on user performance increases as more data becomes available over time, as with task type; it already reaches 85.0% after seeing 10% of the data and plateaus at 90.3% accuracy with 90% of the data observed. The ability of the constructed models to make accurate predictions for performance and task type with 10% of the trial data indicates that these factors influence eye gaze patterns from the very beginning of tasks.

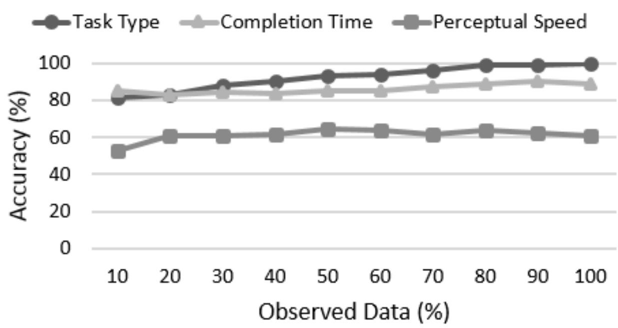

Figure 4. *Trends in classification accuracy over time.*

Predicting user performance and task type is sufficient for providing general adaptations, but to provide support optimized for individual users it's necessary to model user characteristics as well. In contrast to task type and performance, the accuracy of perceptual speed peaks before the end of the interaction. This trend is representative of the trends exhibited by other user characteristics, and is consistent with those reported by Steichen et al. (2013). The trend can be attributed to the diluting effect that generating features over more data has on transient patterns that occur earlier in trials. The peak accuracies for cognitive traits range from 63.5% to 64.5%. Notably, an accuracy within one percent of each of these peak results was obtained in the first half of interactions, indicating that most of the value from gaze features can be extracted while a user is still engaged in a task.

Our obtained accuracies for cognitive traits are better than or comparable to those reported by Steichen et al. (2013), which were obtained using a custom set of AOIs and range between 59% to 64% with a baseline of roughly 50%. This comparison reinforces the finding that using interface-independent features compares favorably to models built with manually specified AOIs. On the other hand, these accuracies remain low for practical use, and while cognitive abilities impact how users execute tasks (Toker et al., 2013), (Steichen et al., 2014b), the difficulty of predicting them compared to task type and performance is an indication that variance in gaze patterns is largely attributable to the nature of the visualization tasks and the way they are executed. Still, our findings imply that gaze data is at least a potential source of information for modeling these user characteristics, which should be further investigated in conjunction with other sources and by using different machine learning approaches (e.g. sequence mining, Steichen et al., 2014b).

Conclusions

This paper investigated the accuracy of predicting user tasks, performance and traits, while users are performing visualization tasks with bar graphs. This research is a step toward user-adaptive visualization systems that can model their users' needs during interaction and provide real-time intervention personalized to satisfy these needs.

We showed that user performance, task type, and four of the user characteristics investigated can be predicted from eye gaze data with accuracies significantly above a majority class baseline, with particularly strong results for task type and performance. These findings mirror results from previous work in which users used bar graphs for solving similar tasks with simpler datasets, thus supporting the robustness of the results to changes in visualization complexity. Furthermore, we showed that using gaze features not customized to the specific interface used in the study delivered comparable accuracies as interface-dependent feature sets. This finding is an encouraging sign that the classification methods discussed in the paper could be generalized across interfaces without requiring the definition of custom features for each one.

Finally, we found that classification accuracy is influenced by the highlighting interventions added to bar graphs to support visualization processing. This influence can be either negative or positive, depending on the intervention and the classification target. Interventions that facilitate continued user and task modeling could be preferred in practice over interventions that are otherwise comparably effective in improving user performance.

As future work, we are planning to investigate other sources of user modeling information from eye-tracking data, including pupil dilation and distance of the user's head from the screen, to augment the classification results obtained with summative eye gaze features alone and to predict additional measures such as user interest and confusion. We also plan to integrate the models and interventions described in this paper into a prototype adaptive system and evaluate its practicality.

References

Jae-wook Ahn and Peter Brusilovsky. 2013. Adaptive visualization for exploratory information retrieval. *Information Processing & Management*, 49(5):1139–1164.

Robert Amar, James Eagan, and John Stasko. 2005. Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the 2005 IEEE Symposium on Information Visualization*, 15–21, Washington, DC, USA. IEEE Computer Society.

Daria Bondareva, Cristina Conati, Reza Feyzi-Behnagh, Jason M. Harley, Roger Azevedo, and François Bouchet. 2013. Inferring Learning from Gaze Data during Interaction with an Environment to Support Self-Regulated Learning. In *Artificial Intelligence in Education*, 7926:229–238. Springer Berlin Heidelberg, Berlin, Heidelberg.

Giuseppe Carenini, Cristina Conati, Enamul Hoque, Ben Steichen, Dereck Toker, and James T. Enns. 2014. Highlighting Interventions and User Differences: Informing Adaptive Information Visualization Support. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 1835–1844. Toronto, Canada.

Cristina Conati and Heather Maclaren. 2008. Exploring the role of individual differences in information visualization. In *Proceedings of the working conference on Advanced visual interfaces*, 199–206, New York, NY, USA. ACM.

Shahram Eivazi and Roman Bednarik. 2011. Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data. In *Proceedings of 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI 2011*, 9–16.

Ruth B. Ekstrom and United States Office of Naval Research. 1996. *Manual for Kit of Factor Referenced Cognitive Tests*. Educational Testing Service.

Stephen Few. 2009. *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press, Oakland, Calif.

Keisuke Fukuda and Edward K. Vogel. 2009. Human Variation in Overriding Attentional Capture. *The Journal of Neuroscience*, 29(27):8726–8733.

David Gotz and Zhen Wen. 2009. Behavior-driven visualization recommendation. In *Proceedings of the 14th international conference on Intelligent user interfaces*, 315–324, New York, NY, USA. ACM.

Beate Grawemeyer. 2006. Evaluation of ERST – An External Representation Selection Tutor. In *Diagrammatic Representation and Inference*, 4045:154–167.

Mark Hall. 1999. *Correlation-based Feature Selection for Machine Learning*. PhD, University of Waikato, April.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

Natasha Jaques, Cristina Conati, Jason M. Harley, and Roger Azevedo. 2014. Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. In *Intelligent Tutoring Systems*, Lecture Notes in Computer Science, 29–38. Springer International Publishing, January.

Marcel Adam Just and Patricia A Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4):441–480.

Samad Kardan and Cristina Conati. 2013. Comparing and Combining Eye Gaze and Interface Actions for Determining User Learning with an Interactive Simulation. In *Proc. of UMAP, 21st Int. Conf. on User Modeling, Adaptation and Personalization*.

Yan Liu, Pei-Yun Hsueh, J. Lai, M. Sangin, M.-A. Nussli, and P. Dillenbourg. 2009. Who is the expert? Analyzing gaze data to predict expertise level in collaborative applications. In *IEEE International Conference on Multimedia and Expo, 2009. ICME 2009*, 898–901.

Mohamed Mouine and Guy Lapalme. 2012. Using Clustering to Personalize Visualization. In *16th International Conference on Information Visualisation (IV)*, 258–263.

Ben Steichen, Giuseppe Carenini, and Cristina Conati. 2013. User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, 317–328, New York, NY, USA. ACM.

Ben Steichen, Cristina Conati, and Giuseppe Carenini. 2014a. Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data. *ACM Transactions on Interactive Intelligent Systems*.

Ben Steichen, Michael M. A. Wu, Dereck Toker, Cristina Conati, and Giuseppe Carenini. 2014b. Te,Te,Hi,Hi: Eye Gaze Sequence Analysis for Informing User-Adaptive Information Visualizations. In *User Modeling, Adaptation, and Personalization*, Lecture Notes in Computer Science, 183–194. Springer International Publishing.

Dereck Toker, Cristina Conati, Giuseppe Carenini, and Mona Haraty. 2012. Towards adaptive information visualization: on the influence of user characteristics. In *Proc. of UMAP, 20th Int. Conf. on User Modeling, Adaptation and Personalization*.

Dereck Toker, Cristina Conati, Ben Steichen, and Giuseppe Carenini. 2013. Individual User Characteristics and Information Visualization: Connecting the Dots Through Eye Tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 295–304, New York, NY, USA.

Dereck Toker, Ben Steichen, Matthew Gingerich, Cristina Conati, and Giuseppe Carenini. 2014. Towards Facilitating User Skill Acquisition - Identifying Untrained Visualization Users through Eye Tracking. In *Proceedings of the 2014 international conference on Intelligent user interfaces*. ACM.

Marilyn L Turner and Randall W Engle. 1989. Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2):127–154.

Maria C. Velez, Deborah Silver, and Marilyn Tremaine. 2005. Understanding Visualization through Spatial Ability Differences. In *IEEE Visualization*, 2005. VIS 05, 511–518.

Caroline Ziemkiewicz, R. Jordan Crouser, Ashley Rye Yauilla, Sara L. Su, William Ribarsky, and Remco Chang. 2011. How Locus of Control Influences Compatibility with Visualization style. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 81–90.