Machine Learning from User Interaction for Visualization and Analytics: A Workshop-Generated Research Agenda

John Wenskovitch,* Michelle Dowling† Laura Grose,‡ Chris North§ Virginia Tech

Remco Chang¶ Tufts University

Alex Endert|| Georgia Tech

David H. Rogers** Los Alamos National Lab

Abstract

At IEEE VIS 2018, we organized the Machine Learning from User Interaction for Visualization and Analytics workshop. The goal of this workshop was to bring together researchers from across the visualization community to discuss how visualization can benefit from machine learning, with a particular interest in learning from user interaction to improve visualization systems. Following the discussion at the workshop, we aggregated and categorized the ideas, questions, and issues raised by participants over the course of the morning. The result of this compilation is the research agenda presented in this work.

Keywords: Machine learning, User interaction, Visualization, Analytics, Research agenda.

Index Terms: Human-centered computing—Visualization

1 Introduction

Interactivity in visualizations and analytical tools provides substantial benefits to the sensemaking processes of users, with domains for such tools ranging from investigative analysis of documents [48] to exploration of scientific simulation results [24]. Recent investigation in the visual analytics domain has begun to explore the benefits of tools and algorithms that learn the intent of a user from their explorations, thereby enabling a system to adapt the layout and contents of a visualization to reflect the user's mental model [4, 22, 26]. As a result, the analytical process can be made more efficient as the system learns from and adapts to the user.

A number of visual analytics tools provide interactive data projections that update based on learned user behaviors [25, 28, 30, 34, 52]. For example, Andromeda [43, 44] and Dis-Function [5] support exploratory dimension reduction for high-dimensional quantitative data. Scientific visualization introduces opportunities for learning from interactions, such as analysis of data ensembles [31], interactive visual querying of nonlinear solution spaces [11], in-situ analysis of large scale data, and data foraging through extreme scale data [51].

This area of research at the intersection of visualization and machine learning is still novel, and much research remains to be done in a number of areas. In this work, we present a research agenda to advance this field, generated through discussions by a group of experts in this domain during the Machine Learning from User Interaction for Visualization and Analytics workshop at IEEE VIS 2018. We discuss challenges and opportunities where further research is needed, including needs for developer creativity to generate novel tools and user studies to better understand the effects of design decisions. This discussion is accompanied by descriptions of state-of-the-art tools that support research in this space.

*email: jw87@vt.edu
†email: dowlingm@vt.edu
‡email: laurag98@vt.edu
§email: north@cs.vt.edu
¶email: remco@cs.tufts.edu
||email: endert@gatech.edu
**email: dhr@lanl.gov

1.1 Methodology

The content that coalesced into this research agenda was generated by participants at the Machine Learning from User Interaction for Visualization and Analytics workshop at IEEE VIS 2018. The high-level goal of this workshop was to bring together researchers from across the visualization community to discuss how machine learning can be used to support visualization tools and workflows. While many systems enable user interaction to explore data and parameter spaces in analytics, this workshop examined how systems can learn from user interactions to iteratively produce even more insightful results. For example, beyond directly manipulating analytical input parameters, users might interact with analytical outputs to better direct the analysis congruent with their interest. This implies a need to learn from the user interactions to transform them into alterations of the parameters for the underlying analytics. In this manner, learning from user interaction can aid sensemaking and performance in analysis [18, 42].

At this workshop, two sessions were designed to include a set of motivating papers, followed by dedicated time for discussion in breakout groups. Approximately 50 attendees participated in these discussion sessions. Discussion groups were semi-supervised by the workshop organizers, and seeded with the discussion goal of identifying applications and open research questions related to the workshop topic. Each group included a note taker, who summarized the group discussion in real-time using a shared Google document.

Following the workshop, the organizers collected these distributed notes into a single document in bullet point form. Within this overarching document, we began an affinity diagramming process to uncover common themes, research topics, and open questions, among other discussion points. This affinity diagramming process was iterative, and we continued to refine topics and discussions in weekly meetings over a period of several months. The topics that resulted from these discussions (found in Sections 2-6) were then placed in a sequence that represents computational and user-driven components in an exemplary human-in-the-loop analytical system framework.

1.2 Human-in-the-Loop Organizational Structure

We have categorized this research agenda into five phases, which interconnect in a hierarchical, cyclical structure reminiscent of Pirolli and Card's Sensemaking Process [37]. Our goal with this organizational structure is to capture the human-in-the-loop process (Fig. 1), bringing forward relevant issues and related work at each step.

We select user interaction as the starting point of our discussion (Sec. 2), which is logged and recorded by the system (Sec. 3). This process can occur iteratively with one or more interactions, until a learning process is initiated (Sec. 4). Following this learning, the new state of the machine learning (ML) model is communicated to (Sec. 5) and interpreted by (Sec. 6) the user, which acts as a second iterative process as the user works to understand the new model. After evaluating the results of their previous interaction(s), the process can begin again with a new set of interactions.

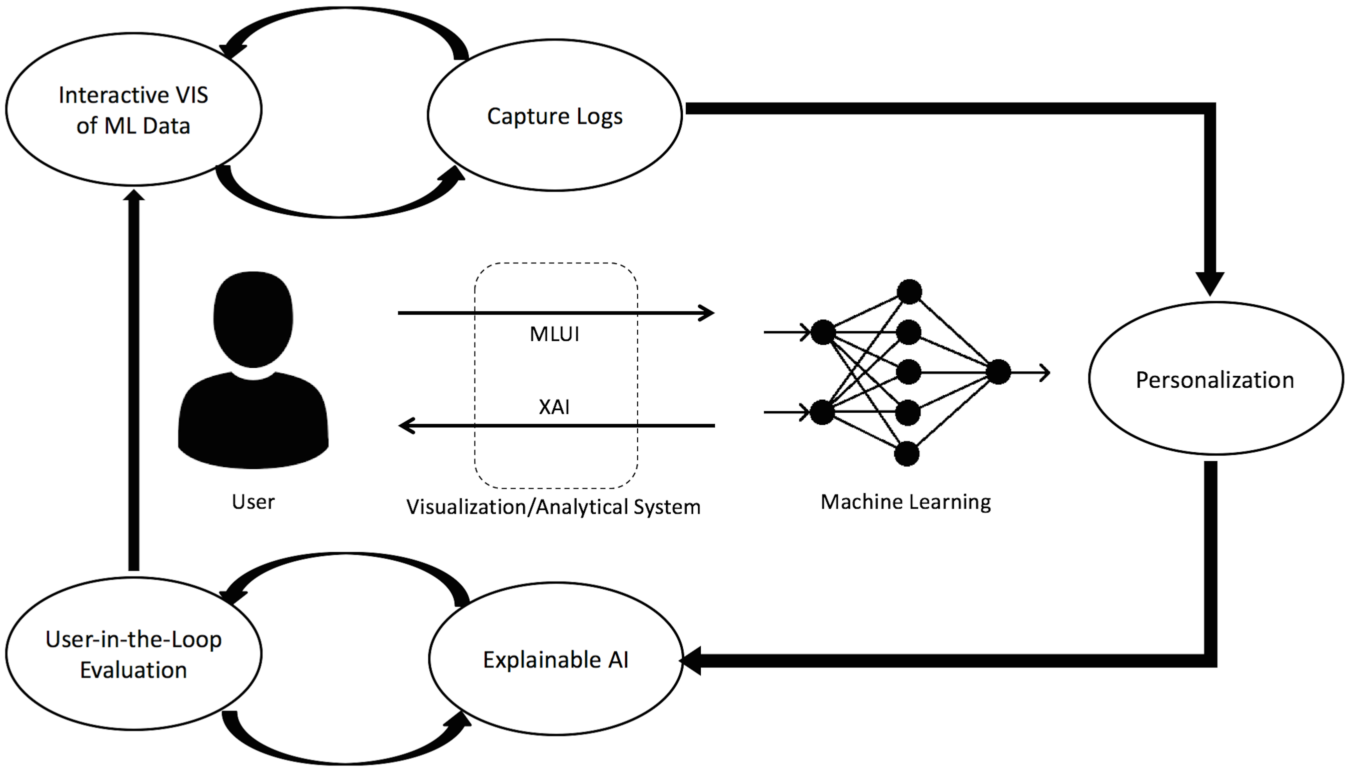

Figure 1: This research agenda captures five interconnected phases represented by each of the ovals: a user interacting with a system (Interactive Visualization (VIS) of Machine Learning (ML) Data), the system logging those interactions (Capture Logs), the system learning from those interactions (Personalization), the system communicating its learned response to the user (Explainable AI), and the user interpreting that learned response (User-in-the-Loop Evaluation). Each of these phases can be roughly categorized by whether the user performs the given phase, the visualization/analytical system, or machine learning within the system. This is division is denoted by the relative position of each of the phases in the central portion of this figure to the different phases. Each phase is explained in more detail in Sections 2-6.

2 Interactive Visualization of Data

We begin our human-in-the-loop process discussion at the point where the user is about to interact with the system. The system has already constructed a visualization of a dataset for the user to explore, and the user has already evaluated the current visualization and determined how well it supports their analysis goals. This visualization could represent either the initial state of the system, or it could be at some later stage within the analysis process. In this section, we discuss how the user understands the interaction techniques available to them and determines which best supports their goal. Our discussion focuses on three separate issues: how should systems guide user interaction, what descriptions are available to allow the user to interpret these interactions, and where does the user even start? This user-centric understanding is tightly coupled with the system capturing interactions to continue learning about the user, as discussed in more detail in the next section. These two phases form an iterative process in which the user may perform multiple logged interactions in order to reach a specific analytical goal.

2.1 Guiding Users Towards Interactions

To help the user understand what they can or perhaps should interact with, machine learning can be used to help guide the user towards an interaction. Two major research questions related to system guidance are **How should systems provide this guidance?** and **How will users interpret the effect of a recommended interaction?**

One form of system guidance is making use of recommendation algorithms to highlight information of interest to an analyst, thereby assisting them in focusing their exploration on relevant information. When considering how to accomplish such recommendations, an important question is, **Which interactions and data should users be guided towards?** This question remains a significant research challenge in which the answer largely depends on the system's perceived importance of data. The idea of information scent and scented widgets [7–9, 36, 53] embeds navigational cues and interaction effects into potential user actions. Such techniques could guide users towards interesting or unique data, recommend commonly-used interactions, or highlight attribute values that have not yet been explored, thereby resulting in additional insight on the data. Further, the system itself could request information from the user directly via an active learning approach, guiding the user towards interactions that the system believes will best improve the model [45].

A related question is, **What visual metaphor should the system use to convey this guidance?** For one example, StarSPIRE [4] and Cosmos [17] recommend documents (i.e., data to explore) to the user by filtering documents to display based on the perceived relevance of those documents to the user's current analysis. This perceived relevance is mapped to the corresponding node sizes of the displayed documents, providing the user with more information regarding this recommendation that denotes which of those documents may be most relevant to them. Alternatively, ModelSpace [6] (see Fig. 2) shows a projection of the system states that an analyst has already explored, providing a visual reminder of unexplored regions that should be evaluated for completeness.

Related to the previous questions, **How should the system guide user interactions in such a way that analysts will understand the potential effects of an interaction?** ML responses are complex and difficult to predict, which might cause fear or uncertainty

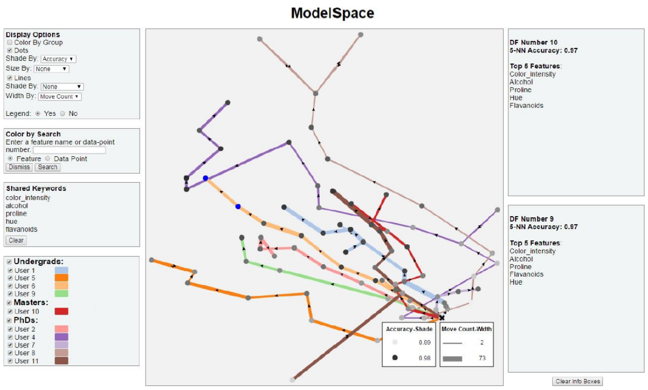

Figure 2: A screenshot of ModelSpace [6], which visualizes the different states that have been explored by each user.

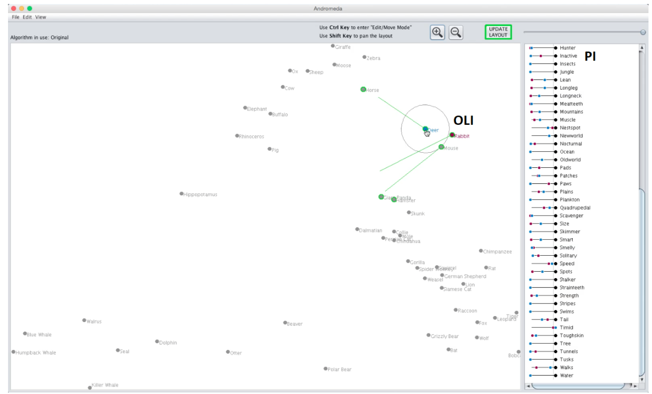

Figure 3: Andromeda supports parametric interaction (PI) with interactive slider widgets and observation-level interaction (OLI) with direct manipulations in the projection.

of interaction should be performed. For example, users who are unfamiliar with complicated ML algorithms that are being used to create the visualization may believe an interaction will have a small impact. When the interaction results in a much larger change, the user may be confused since their mental mapping for the effects of the interaction was different than the results show. An example of a visual metaphor a system can use to guide interactions that may reduce this confusion is providing a preview of the resulting changes to the visualization after the interaction. Such previews help the user decide if the given interaction will produce desired results. Similarly, the ability to undo an action may provide a solution though which the user can return to a previous state if the interaction produces undesired results. The ability to undo an action is also critical. The Speculative Execution model proposed by Sperrle et al. [47] presents another solution to this challenge, computing potential future model states and presenting them in comparison to the current system state. This permits analysts and domain experts to see the effects of interactions before they are initiated (see Fig. 4).

2.2 Matching User and System Understandings of Interactions

When a user misunderstands the interactions available to them, they can become frustrated, misuse the system, follow bad exploration paths, or reach incorrect conclusions. Therefore, it is important to understand **How can we mitigate users' misunderstandings of the available interactions and their effects?** A proposed solution is to generate standardized interaction terminology to provide the user with an understandable mapping between each interaction and the system's reaction to performing that interaction.

However, to provide such terminology, it is important to know **How and when do user misunderstandings of interactions occur?** Using a simple menu system as an example, misunderstandings and frustration can occur if a menu item is poorly labeled, leaving the user to guess or misinterpret what the menu item does. If instead the menu is well-designed and follows established conventions, like "Save" always being under a menu category called "File," users can easily navigate and use the menu system effectively.

Once how and when users' misunderstandings of interactions can occur are better understood, this knowledge provides some insight into the users' mental mapping of interactions to system responses and updates. Therefore, an important consideration is, **How should the system provide an interaction (i.e., how should an interaction be designed) such that the user learns how the system maps the interaction to its influence on the visualization?**

For example, in Andromeda [42] (Fig. 3), parametric interactions (PI) are mapped closely to model parameters. This mapping is expressed to the user by displaying and manipulating attribute weights with interactive sliders, an intuitive interaction familiar to users. In contrast, Andromeda maps observation-level interactions (OLI) to direct manipulations of the projected observations. Ideally, the user's understanding of this interaction should be that such direct manipulations express desired similarity and dissimilarity relationships between observations. However, this mental mapping can be more difficult for users to grasp, in part because the mapping between the interaction and model parameters is less clear in OLI. However, another reason may simply be that users are not as accustomed to directly interacting with projections of data as they are with sliders. In other words, users perceive sliders as having a higher interaction affordance than projected observations.

2.3 System Starting Point

Thus far, we have assumed that the user has a visualization with relevant data displayed for them to interact with. However, this may not be the case, especially when provided with an initial visualization to begin the data exploration process. Indeed, because the view that is provided to a user will anchor them in their current and future sensemaking interactions, an important research question is, **Which initial visualization is most appropriate for the given user tasks and data?** This question becomes more complex is a system provides multiple views or perspectives of the data.

For example, systems such as StarSPIRE [4] and Cosmos [17] begin with no data displayed; users must explicitly search for data to begin seeing anything on the screen. This means that the visualization will not bias the user towards any data. However, the tradeoff is that the user must have some idea of what they want to begin exploring in their data. Alternatively, systems like Andromeda [43] display all available data in the initial visualization. An advantage of displaying all the data initially is that users can get an overview of the entire dataset, allowing them to gain a high-level understanding from the beginning. The tradeoff is that such visualizations do not scale as well to larger datasets. However, there is also the potential to explore initial visualizations between these two extremes. Additionally, more complex systems that include multiple visualizations, such as dashboards, exacerbate this issue since how to initialize each individual visualization will have to be determined.

3 Capture Logs

The interactions performed by a user indicate their current intent or goals in their analysis. Therefore, logging these task-oriented interactions can help the system disambiguate these intents and goals to further assist the analyst in achieving said goals. In this section, we discuss the challenges of deciding what to log, provenance of

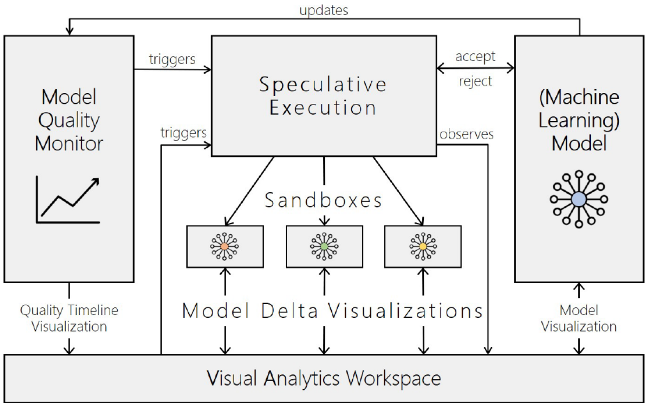

Figure 4: The Speculative Execution concept from Sperrle et al [47].

logs, and the importance of context. These logs are used in the next phases to update models and personalize visualizations.

3.1 The Art of Logging

Ultimately, logs are how the machine is able to capture the user's progress and forms the foundation for understanding user intent. Therefore, ensuring that logs are being captured in a manner that enables the personalization (accomplished in the next step) based on this learned knowledge is important. A critical question to ask is, **How can we determine what the important interactions are?** Logging every pixel of mouse movement and timing every fractional pause is certainly an unsustainable scale of data but guarantees interaction coverage. In contrast, determining that an interaction happened by waiting for a purposeful action such as a mouse click or keystroke may miss important contextual information for that action.

A structured means of considering this challenge which still has inherent open questions is to determine **What to log, when to log, and how to log?** "What to log" focuses on the type of interactions being performed (e.g., clicks and keystrokes); not every interaction may be important (e.g., hovering to see a tooltip). In contrast, "when to log" refers to when such logs should be generated (e.g., only tracking the first click on an object rather than all clicks). These two considerations relate to a notion of scalability of the logs; capturing every interaction means that the logs will become difficult to analyze efficiently or effectively. Lastly, "how to log" centers on media to gather or generate logs, such as eye tracking or audio recording. These considerations dictate what information the system ultimately is able to use in the personalization step and, by extension, how the information can be used.

However, a single interaction does not convey much information; a sequence of interactions can convey much more about the user's intent and their current goal. Therefore, being able to track interactions over time is another important consideration. This points to a notion of provenance of interactions and maintaining provenance in the logs [15, 38], and leads to another open question: **How can a generic system retrieve high-level interactions from a sequence of lower-level interactions?** For example, the system can use provenance data to learn which interactions are important or preferred. With this information, the system can keep more detailed logs about the important interactions [21], which can help dictate what can be or should be learned from such interactions.

3.2 Contextualizing the Interaction

The idea of provenance leads to considering context in the interaction. **What is the user interacting with? What data is currently being used or considered? What is the user's current state in their analysis process?** These different facets of context help situate the interaction within the user's analysis process. Thus, context can enable the system to personalize the visualization in the next update based on the user's current process as opposed to only the user's current state. This information about the user's process may include information such as what subset of data is most relevant to the user and the scope of the user's current process. In this sense, logging (particularly with context) opens the "black box" of the user to help the system better understand the user's analysis process and goals [35].

To assist with identifying context, the interactions can be taxonomized based on how or when they are used and what data they are being used with. Such a taxonomy would help identify what contextual information should be captured alongside the interactions. If a notion of context for an interaction can be predefined, then the system can also begin determining what the user's intent behind using that interaction may be. Existing interaction taxonomies [3, 46, 50, 54] would benefit from contextual extensions.

4 Personalization

With the interaction logs, the system can attempt to infer the user's intent and provide an updated, personalized visualization to help with the user's analysis. This personalization can focus on user characteristics, like personality and experience level, as well as their intent. However, such personalization can be difficult to achieve depending on the information that is logged and how indicative it is towards these personalization goals. The system can communicate this personalized new state back to the user, which will be described in the next section.

4.1 Predicting User Intent

Predicting the intent of the user will help the system provide a visualization that is tailored to the user's goals, but predicting their intent is complicated. The user may also have multiple, parallel analytical goals when exploring their data. Even if the user has a single or primary goal, they will likely perform multiple interactions to achieve that goal. **How can a system deal with this cardinality issue, appropriately mapping interaction sequences to the goals of the user?** In one proposed solution by Dowling et al. [16], the system consists of a computational pipeline that maps specific subsets of interactions to individual models the pipeline (see Fig. 5 for an example). This provides a structure to begin mapping user interactions to overarching goals.

If the system had advance knowledge of which interactions denote which user intents, it could develop a taxonomy or grouping of such interactions to assist in learning these intents. Additionally, the system can use a series of interactions to gain more information about the user's intent, making techniques such as human-in-the-loop analytics particularly powerful since such techniques require regular feedback from the user.

Semantic interaction takes this idea a step further. When performing semantic interactions, the user is shielded from the details of the underlying models, but as a result, it may be less clear how their interactions influence the model and thereby influence future visualizations. Though a number of systems that incorporate semantic interactions have been implemented, it is still unclear **What is the optimal means of translating semantic interactions into model updates?** These systems run the gamut from making purely heuristic updates to solving equations to determine precise parameter updates. Understanding the underlying cognitive state of the user can give some clues as to the intent of the user [1].

4.2 Personalization for User Personality

An alternative perspective on personalization is to utilize the user's personality traits. These personality traits can help further refine the system's response to user interactions to provide an improved

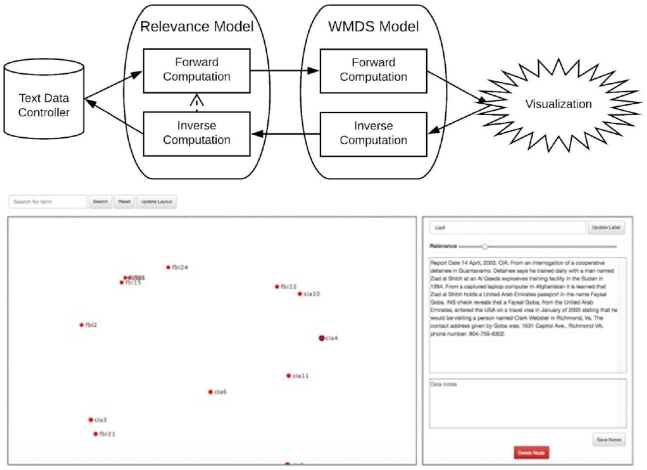

Figure 5: An example computational pipeline and system from Dowling et al [16], in which interactions with document relevance are handled by the Relevance Model and interactions with document positioning are handled by the WMDS Model.

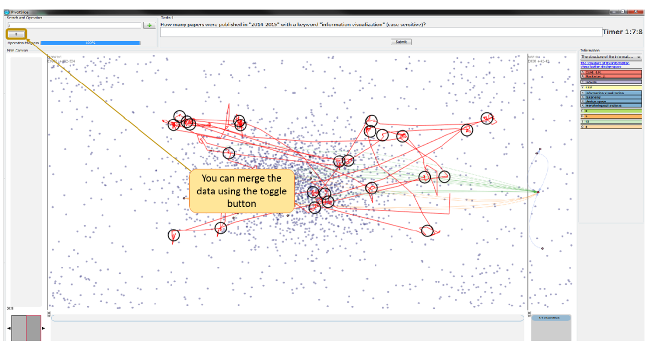

Figure 6: After a system detects frustration, it can display suggested interactions [33].

visualization and overall experience. For example, if the system knows that the user is confused or frustrated [33], perhaps it can guide the user toward a useful interaction or relevant data. An example of such a suggested interaction is demonstrated in Fig. 6. However, frustration is a short-term trait that can change with time; other traits like expertise are longer-term. Knowing which traits are long-term vs. short-term, **How can a system learn these traits for an individual user?** Incorporating such knowledge in the system's personalization is an active challenge and open research direction. Additionally, **How can user personality characteristics best be used to create better and more useful visualizations and user experiences?**

Further, personality traits may be characterized by different types of users [20, 32, 41, 55]. For example, the system could leverage a grammar of interactions for model tuning to assist expert users (e.g., model builders) in their goals. This grammar may be different for other types of users, such as domain experts or managers.

4.3 Information that is Used / Should be Used

To accomplish the aforementioned goals for personalization, **What information do we need from logs to accomplish personalization goals?** Aboufoul et al. propose a method to identify the cognitive processes of users based upon logged interactions [1] (Fig. 7). The previously captured logs of interactions can help map user intents to higher-level goals, but the system needs a taxonomy of interactions

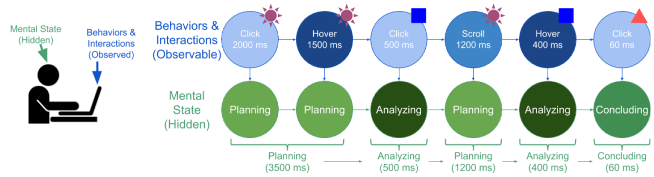

Figure 7: Using user behaviors and interactions to predict corresponding hidden mental states [1].

and intents. Such a taxonomy can therefore assist in linking interactions and intents to manipulations of machine learning parameters.

The type of information that is logged directly affects the system's ability to infer from such a taxonomy. Here, we see the role of context mentioned in Section 3.2 take effect. For example, the timing of these interactions can provide insight into the user's current thought process or state in analysis. Additionally, if the user is focused on a subset of data, the system could take this context into account as it determines how best to personalize an updated visualization for the user. Thus, by taking advantage of many sources of information, the system can create a personalized updated visualization.

5 Explainable AI (XAI)

After the system has updated and personalized the underlying models, the system should now provide an updated visualization and corresponding explanation of the current state of underlying models. Ideally, the system providing this feedback will enhance user trust by opening the "black box" of machine learning, thereby enabling the user to understand how the system reached this state. In this section, we discuss the difficulty of providing understandable explanations to the user. Once the user has received this feedback, they can go on to properly evaluate these changes and their suitability in their current analysis process. The communication from the system and the interpretation from the user is a tightly-coupled and iterative process, similar to that between the Interactive Visualization and Capture Logs phases (see Fig. 1). Another IEEE VIS workshop, VISxAI[1], was dedicated to Explainable AI. As such, there was limited discussion at the MLUI 2018 workshop on this topic.

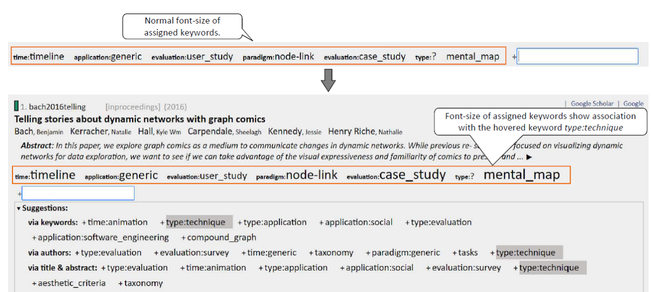

Figure 8: The relationship between suggested and assigned keywords encoded in font size.

5.1 Providing Feedback to the User

This step of the loop focuses on how to open the "black box" of the underlying machine learning algorithms to help the user better understand the updates that the system has performed. Thus, this step is equivalent to explainable artificial intelligence (XAI). Given how rich this area of research is [13, 14, 19, 29, 39], we wish to simply focus on components of XAI that are relevant to the user's

[1] `http://visxai.io/`

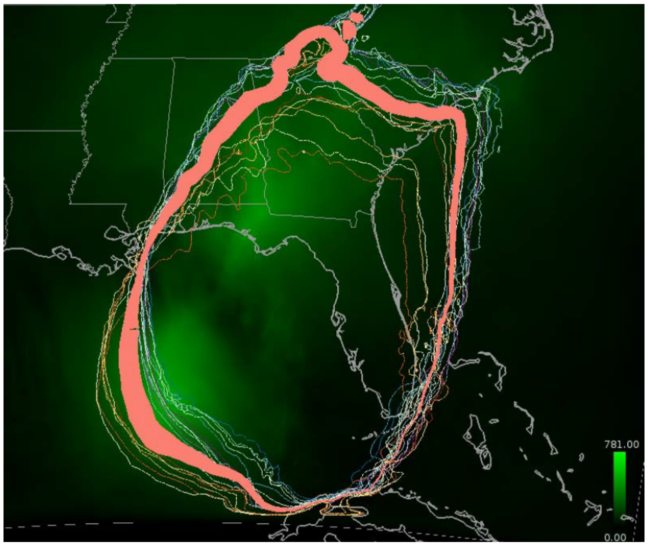

Figure 9: An uncertainty ribbon showing the interquartile range of perturbation pressure in a weather model simulation ensemble [40].

perspective on this step of the process. In particular, an open question is, **What types of feedback will benefit the user most?** For example, feedback can be given in terms of how specific model parameters changed (requiring expert knowledge of the model itself), or feedback can be provided in more natural and intuitive manners to the user (which may obscure details regarding how specific model parameters have changed). As a result, two major considerations in how to provide this feedback to the user are (1) how simple the explanations should be and (2) how much background knowledge is required to understand the explanations. For example, the literature tagging application from Agarwal et al. [2] demonstrates a straightforward visual explanation of relationships between suggested and assigned keywords through font size and hovering (Fig. 8).

5.2 Interpretability and Uncertainty

Related to these major considerations are two other facets of XAI: interpretability and uncertainty. These facets of XAI focus on not just how the user can understand the high-level changes that occurred in the underlying models but also the implications and accuracy of these changes. For example, in many dimension reduction techniques, several similar projections may be produced with the same parameters. If the user can understand this uncertainty in the projection, perhaps they can better understand more nuanced relationships in the data, thereby creating a more detailed mental model of the data that further aids their analysis process. However, displaying information regarding uncertainty in a manner that is understandable or interpretable to the user is a challenging task. As such, a number of uncertainty visualization mechanisms have been designed, including the uncertainty ribbons created by Sanyal et al [40] (Fig 9). Indeed, **How do users even interpret uncertainty?** A system that communicates uncertainty in its output could be interpreted on a scale that ranges between open and honest with its current knowledge and useless because it will not provide a precise answer. Similar ideas centered on the interpretability of the machine learning feedback are also a growing area of research [10, 23, 49].

6 User-in-the-Loop Evaluation

After the user has received feedback from the system regarding the updated models, they can begin to evaluate the updated visualization and its suitability in their current analysis process. This evaluation process is tightly coupled with the feedback provided by the system

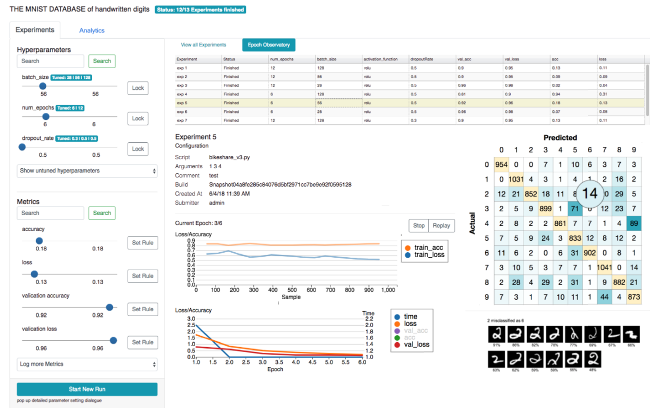

Figure 10: The experiment dashboard in HyperTuner [27].

and requires one or more metrics, which may be chosen by either the system or the user. Choosing the correct metrics and how to integrate them into human-in-the-loop systems is a challenge; these metrics are ultimately determined by the user based on their current goal, which may be difficult to externalize in an explicit or easily measurable manner. After performing this evaluation on the updated visualization's suitability to their task, the user can interact with the visualization once again to continue their analysis process and provide additional feedback to the system.

6.1 How to Evaluate

The goal of evaluating the system is for the user to determine the suitability of the updated visualization in their current analysis process. **How should a user properly evaluate the solutions that are presented by the system?** This evaluation will be influenced by the feedback provided to the user in the previous step; however, this evaluation will be based on user-defined metrics from their own mental model for the visualization's suitability. To support this evaluation, the system should be designed to provide the evaluation metrics that are more informative to the users. For example, if the user's various tasks mandate fast responses from the underlying model, then the system could be built to provide a measurement of how much performance could increase based on which subsets of the data are used in the subsequent calculation (i.e., after the next interaction). For example, HyperTuner [27] gives analysts the ability to perform a hyperparameter search interactively, evaluating a collection of models in a sequence of experiments and passing the results of predefined metrics back to the user (see Fig. 10). Further, Corbett et al. [12] present a set of ten heuristics for evaluating interactive machine learning systems.

In a similar vein, **How should the user and the system handle bias?** If the user's higher-level tasks or goals are concerned with bias, the system may be built to provide metrics of bias as part of its feedback. Alternatively, it may learn (through user interaction) that bias is important to the user in their current evaluation and therefore *learns* to provide metrics for bias. Similarly, users can provide information on past models that they evaluated to be useful. Therefore, the system can learn about its visualization's suitability in the user's current analysis process through user interaction in the next step (and subsequently logging and personalization).

6.2 Trust

Having a proper evaluation of the system may aid the system in establishing trust with the user. But, **How does a user decide to trust what they see?** How the user evaluates the updated visualization is reliant on how much the user trusts the update and associated feedback from the previous step. To help the user trust the system more, the system can learn what a user needs to trust the visualization and feedback. Similarly, it is important for the system to know or learn how and when breakdowns of trust occur to further guide how it provides feedback to the user or personalizes visualizations. Thus, a related questions is, **Can a system learn what users need in order to trust a system?** In other words, trust may change over time, meaning the system may have to re-evaluate the user's trust as the system progresses from one state to the next. This is a complex and open research direction.

7 Discussion

7.1 System Self-Correction, or How to Overcome Incorrect Inferences

One concept that spans multiple steps in our process is the idea of a system self-correcting or overcoming incorrect inferences. This concept targets the fact that trying to learn user intent (as is done in the personalization step) may lead to incorrect conclusions. The result of such incorrect conclusions is that the user will be provided with an updated visualization that they evaluate as not being suitable in their current analysis process. The system should find ways to correct for these errors.

A method for this correction that we have already discussed is to continue to try learning the intent from users based on their logged interactions. However, how different might the updated visualization be if the system assumed that it was likely wrong in trying to infer the user's intent? When the system learns that it was, in fact, wrong, does it understand why its models failed to then know how to correct for this error? For example, how does the system compensate for the fact that it can understand relationships in the data (like correlation), but users can understand other relationships as well (such as causal relationships)? From a slightly different perspective, what is the difference between a system knowing how to self-correct and a user telling it to correct (through interaction)?

The optimal method to perform this correction is largely based on the specific user in question. For example, expert users are more likely to recognize and understand when the system is providing a nonsensical visualization, whereas novice users may not. Similarly, expert users can give more detailed feedback, such as what the boundaries of the models should be and what the edge cases are. Therefore, it is important for the system understand the type of user and their personality as well as the data itself and which subsets of the data the user is interested in.

While it is certainly important to look at examples of systems that have a notion of self-correction, it is also useful to look at examples of systems that do this poorly. A frequently-cited example is Microsoft's Clippy. Clippy has become known as an example of what an AI should *not* be. The graphic popped up and interrupted the user frequently, many times with incorrect assumptions about what the user was currently trying to accomplish. Furthermore, it never retained knowledge of whether it was right or wrong in trying to learn the user's intent, meaning Clippy never tried to learn user intent differently or did any self-correction. Such mistakes, while perhaps impossible to avoid entirely, should be understood so that any issues the user encounters as a result can be mitigated.

7.2 Future Work / Limitations

The ideas presented here are a reflection of the research and discussion that happened during the IEEE VIS 2018 Workshop on Machine Learning from User Interaction for Visualization and Analytics. As opposed to a thorough literature exploration of the phases in the human-in-the-loop process (Fig. 1), we instead present an overview of the common themes and ideas that arose from the workshop. Additionally, we identify and discuss current research questions and considerations of researchers at the workshop. Thus, a thorough survey of interactive machine learning, human-in-the-loop analytic systems, and human-centered machine learning literature could uncover issues and open questions is not discussed in this work, and we anticipate that the future research directions in these areas will be much richer than the set of ideas presented here.

8 Conclusion

Machine learning from user interaction can offer many new opportunities for visualization and analytics, as demonstrated by the number of open research questions. As such, much work remains to be accomplished in this space, and initial research successes demonstrate encouraging results. Overall, the workshop discussants came to the clear conclusion that the visualization research community should continue with future MLUI workshops and other publication, outreach, education, funding, and community building initiatives for this topic.

Acknowledgments

Thanks to all who submitted to and/or attended the MLUI 2018 workshop.

References

[1] M. Aboufoul, R. Wesslen, I. Cho, W. Dou, and S. Shaikh. Using hidden markov models to determine cognitive states of visual analytic users. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[2] S. Agarwal, J. Bernard, and F. Beck. Computer-supported interactive assignment of keywords for literature collections. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[3] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 111–117, Oct 2005. doi: 10.1109/INFVIS.2005.1532136

[4] L. Bradel, C. North, L. House, and S. Leman. Multi-model semantic interaction for text analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 163–172, Oct 2014. doi: 10.1109/VAST.2014.7042492

[5] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 83–92, Oct 2012. doi: 10.1109/VAST.2012.6400486

[6] E. T. Brown, S. Yarlagadda, K. Cook, R. Chang, and A. Endert. Modelspace: Visualizing the trails of data models in visual analytics systems. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[7] E. H. Chi, L. Hong, M. Gumbrecht, and S. K. Card. Scenthighlights: Highlighting conceptually-related sentences during reading. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, IUI '05, pp. 272–274. ACM, New York, NY, USA, 2005. doi: 10.1145/1040830.1040895

[8] E. H. Chi, L. Hong, J. Heiser, and S. K. Card. Scentindex: Conceptually reorganizing subject indexes for reading. In *2006 IEEE Symposium On Visual Analytics Science And Technology*, pp. 159–166, Oct 2006. doi: 10.1109/VAST.2006.261418

[9] E. H. Chi, L. Hong, J. Heiser, S. K. Card, and M. Gumbrecht. Scentindex and scenthighlights: productive reading techniques for conceptually reorganizing subject indexes and highlighting passages. *Information Visualization*, 6(1):32–47, 2007.

[10] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Gated feedback recurrent neural networks. In *International Conference on Machine Learning*, pp. 2067–2075, 2015.

[11] D. Coffey, C. Lin, A. G. Erdman, and D. F. Keefe. Design by dragging: An interface for creative forward and inverse design with simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2783–2791, Dec 2013. doi: 10.1109/TVCG.2013.147

[12] E. Corbett, N. Saul, and M. Pirrung. Interactive machine learning heuristics. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[13] M. G. Core, H. C. Lane, M. Van Lent, D. Gomboc, S. Solomon, and M. Rosenberg. Building explainable artificial intelligence systems. In *AAAI*, pp. 1766–1773, 2006.

[14] F. K. Došilović, M. Brčić, and N. Hlupić. Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp. 0210–0215. IEEE, 2018.

[15] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *IEEE Computer Graphics and Applications*, 29(3):52–61, May 2009. doi: 10.1109/MCG.2009.49

[16] M. Dowling, J. Wenskovitch, P. Hauck, A. Binford, N. Polys, and C. North. A bidirectional pipeline for semantic interaction. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[17] M. Dowling, N. Wycoff, B. Mayer, J. Wenskovitch, S. Leman, L. House, N. Polys, C. North, and P. Hauck. Interactive visual analytics for sensemaking with big text. *Big Data Research*, 16:49 – 58, 7 2019. doi: 10.1016/j.bdr.2019.04.003

[18] A. Endert, P. Fiaux, and C. North. Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, Dec 2012. doi: 10.1109/TVCG.2012.260

[19] D. Gunning. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*, 2, 2017.

[20] R. Hastie and P. A. Kumar. Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, 37(1):25, 1979.

[21] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189–1196, Nov 2008. doi: 10.1109/TVCG.2008.137

[22] L. House, S. Leman, and C. Han. Bayesian visual analytics: Bava. *Statistical Analysis and Data Mining*, 8(1):1–13, 2015. doi: 10.1002/sam.11253

[23] T. R. Jones, A. E. Carpenter, M. R. Lamprecht, J. Moffat, S. J. Silver, J. K. Grenier, A. B. Castoreno, U. S. Eggert, D. E. Root, P. Golland, et al. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proceedings of the National Academy of Sciences*, 106(6):1826–1831, 2009.

[24] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual analytics: Scope and challenges. In *Visual data mining*, pp. 76–90. Springer, 2008.

[25] B. C. Kwon, H. Kim, E. Wall, J. Choo, H. Park, and A. Endert. Axisketcher: Interactive nonlinear axis mapping of visualizations through user drawings. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):221–230, Jan 2017. doi: 10.1109/TVCG.2016.2598446

[26] S. C. Leman, L. House, D. Maiti, A. Endert, and C. North. Visual to parametric interaction (v2pi). *PLoS ONE*, 8(3):1–12, 03 2013. doi: 10.1371/journal.pone.0050474

[27] T. Li, G. Convertino, W. Wang, H. Most, T. Zajonc, and Y.-H. Tsai. Hypertuner: Visual analytics for hyperparameter tuning by professionals. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[28] G. M. H. Mamani, F. M. Fatore, L. G. Nonato, and F. V. Paulovich. User-driven feature space transformation. *Computer Graphics Forum*, 32(3pt3):291–299, 2013. doi: 10.1111/cgf.12116

[29] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 2018.

[30] V. Molchanov and L. Linsen. Interactive Design of Multidimensional Data Projection Layout. In N. Elmqvist, M. Hlawitschka, and J. Kennedy, eds., *EuroVis - Short Papers*. The Eurographics Association, 2014. doi: 10.2312/eurovisshort.20141152

[31] D. Orban, D. F. Keefe, A. Biswas, J. Ahrens, and D. Rogers. Drag and track: A direct manipulation interface for contextualizing data instances within a continuous parameter space. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):256–266, Jan 2019. doi: 10.1109/TVCG.2018.2865051

[32] A. Ottley, H. Yang, and R. Chang. Personality as a predictor of user strategy: How locus of control affects search strategies on tree visualizations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 3251–3254. ACM, New York, NY, USA, 2015. doi: 10.1145/2702123.2702590

[33] P. Panwar, A. Bradley, and C. Collins. Providing contextual assistance in response to frustration in visual analytics tasks. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[34] F. Paulovich, D. Eler, J. Poco, C. Botha, R. Minghim, and L. Nonato. Piecewise laplacian-based projection for interactive data exploration and organization. *Computer Graphics Forum*, 30(3):1091–1100, 2011. doi: 10.1111/j.1467-8659.2011.01958.x

[35] F. C. Peña and J. Guerra-Gomez. Opening the black-box: Towards more interactive and interpretable machine learning. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[36] P. Pirolli. A theory of information scent. *Human-computer interaction*, 1:213–217, 2003.

[37] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis*, 5:2–4, 2005.

[38] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen. Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):31–40, Jan 2016. doi: 10.1109/TVCG.2015.2467551

[39] W. Samek, T. Wiegand, and K.-R. Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017.

[40] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, Nov 2010. doi: 10.1109/TVCG.2010.181

[41] S. Schwartz. Individual differences in cognition: Some relationships between personality and memory. *Journal of Research in Personality*, 9(3):217–225, 1975.

[42] J. Z. Self, M. Dowling, J. Wenskovitch, I. Crandell, M. Wang, L. House, S. Leman, and C. North. Observation-level and parametric interaction for high-dimensional data analysis. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(2):15:1–15:36, June 2018. doi: 10.1145/3158230

[43] J. Z. Self, X. Hu, L. House, S. Leman, and C. North. Designing usable interactive visual analytics tools for dimension reduction. In *CHI 2016 Workshop on Human-Centered Machine Learning (HCML)*, May 2016.

[44] J. Z. Self, R. Vinayagam, J. T. Fry, and C. North. Bridging the gap between user intention and model parameters for data analytics. In *SIGMOD 2016 Workshop on Human-In-the-Loop Data Analytics (HILDA 2016)*, June 2016.

[45] B. Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.

[46] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*, pp. 364–371. Elsevier, 2003.

[47] F. Sperrle, J. Bernard, M. Sedlmair, D. Keim, and M. El-Assady. Speculative execution for guided visual analytics. In *Proceedings of the Machine Learning from User Interaction for Visualization and Analytics Workshop at IEEE VIS 2018*, Oct 2018.

[48] J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

[49] S. Stumpf, V. Rajaram, L. Li, M. Burnett, T. Dietterich, E. Sullivan, R. Drummond, and J. Herlocker. Toward harnessing user feedback for machine learning. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pp. 82–91. ACM, 2007.

[50] E. R. A. Valiati, M. S. Pimenta, and C. M. D. S. Freitas. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, pp. 1–6. ACM, New York, NY, USA, 2006. doi: 10.1145/1168149.

1168169

[51] C. Ware, D. Rogers, M. Petersen, J. Ahrens, and E. Aygar. Optimizing for visual cognition in high performance scientific computing. *Electronic Imaging*, 2016(16):1–9, 2016.

[52] J. Wenskovitch and C. North. Observation-level interaction with clustering and dimension reduction algorithms. In *Proceedings of the 2Nd Workshop on Human-In-the-Loop Data Analytics*, HILDA'17, pp. 14:1–14:6. ACM, New York, NY, USA, 2017. doi: 10.1145/3077257.3077259

[53] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, Nov 2007. doi: 10.1109/TVCG.2007.70589

[54] J. S. Yi, Y. a. Kang, and J. Stasko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, Nov 2007. doi: 10.1109/TVCG.2007.70515

[55] C. Ziemkiewicz, A. Ottley, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang. How visualization layout relates to locus of control and other personality factors. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1109–1121, July 2013. doi: 10.1109/TVCG.2012.180