Digests, snapshots, events, or cumulative gaze - what is most informative of user success and failure? A study of the foretelling signs of user performance during interaction with visualizations of ontology class hierarchies

Bo Fu
*Computer Engineering and Computer Science, California State University Long Beach, 1250 Bellflower Blvd, Long Beach, CA, 90840, USA*

**Abstract.** The current research landscape in ontology visualization has largely focused on tool development yielding an extensive array of visualization tools. Although many existing solutions provide multiple ontology visualization layouts, there is limited research in adapting to an individual user’s performance, despite successful applications of adaptive technologies in related fields including information visualization. In an effort to innovate beyond traditional one-size-fits-all ontology visualizations, this paper contributes one step towards realizing user adaptive ontology visualization in the future by recognizing timely moments where users may potentially need intervention, as real-time adaptation can only occur if it is possible to correctly predict user success and failure during an interaction in the first place. Building on a wealth of research in eye tracking, this paper compares four approaches to predictive gaze analytics through a series of experiments that utilizes scheduled gaze digests, irregular gaze events, last known gaze status, as well as all gaze captured for a user at a given moment in time. Experimental results suggest that irregular gaze events are most informative of early predictions, while increased gaze is most often associated with peak accuracies. Furthermore, cognitive workload appears to be most indicative of overall user performance, while task type may impact predictive outcome irrespective of the gaze analysis approach in use.

**Keywords:** Ontology Visualization, Gaze Analytics, User Success and Failure Prediction, Indented List, Node-Link Diagram

1. **Introduction**

The benefits of imageries as well as the importance of graphical and visual representations of data and information in human decision-making have been studied and reviewed extensively [62]. Visualizing complex data models such as ontologies has remained a key area of research in the semantic web community, where ontology visualization has become a cornerstone in knowledge editing and discovery tools. While significant research efforts have focused on developing various visualization techniques and layouts for ontologies as surveyed in [12, 13], one observation is the dominance of a one-size-fits-all approach across existing ontology visualization tools. In commonly used visualization techniques for ontologies, users are typically presented with the same (suite of) visualization and interactive support,
which often ignore differences in the individual user needs. There is a wealth of research [63-66] in the field of information visualization (InfoViz) on individual differences in user characteristics, such as cognitive abilities, perceptual speed, working memory, and how such differences in users may impact on their visual search behaviors and overall performances. To account such differences in users, recent research efforts have investigated potential solutions towards user adaptive visualizations depending on user task [34], expertise [35], and monitoring [36, 46], with results showing improved user performance such as task speed and user success.

In the field of ontology visualization, differences in users are typically recognized as preferences for certain types of visualization or specific visual layouts over alternatives. Tools such as [30] address such differing user preferences by providing multiple coordinated views of the same ontological data models that complement one another. In the scope of one ontology visualization, there has been limited research to tailor its visual cues and components within to adapt to individual users. Motivated by the success of user adaptive visualization systems such as [67], whereby adaptive overlays are added to visualizations to aid chart reading, this paper aims to investigate potential means to detect user needs for intervention as the first step towards the realization of adaptive ontology visualizations. In other words, only when it is possible to recognize those moments where a user appears to struggle with an ontology visualization, can we provide personalized visual assistance for that person.

To inform the adaptation process, typical strategies in InfoViz have explored the use of task properties, user expertise, direct user input such as mouse clicks, as well as indirect user input such as eye gaze [34-39]. In ontology visualization, the feasibility of gaze as an input to predict user success and failure has been demonstrated in [59-61] with experimental results showing accurate predictions both in real time and post interaction in the presence of mixed user backgrounds and task domains across commonly used techniques. Given the streaming nature of gaze data, the technique to analyze this data may impact on the subsequent predictions based upon them. To this end, this paper explores four different approaches that analyze gaze as non-overlapping data chunks that are either scheduled or event-based, as well as data chunks that either overlap in part or entirely. Specifically, the goal of this paper is to compare four approaches to gaze analytics, namely gaze digests (non-overlapping scheduled data), gaze events (non-overlapping and non-scheduled data based on significant gaze events detected during interaction), gaze snapshots (partly overlapping data capturing the most recent state of user gaze), and cumulative gaze (entirely overlapping data capturing all known gaze of a user) in a series of experiments to quantify potential impact of these approaches on the predictions of user success and failure for adaptive ontology visualizations.

It is envisioned that these predictions will in turn enable adaptations of future ontology visualization systems with the overall goal of improved user performance and experience during human-ontology interaction. Prior research such as [44] has investigated user predictions in the context of whether the user is likely to succeed using one type of visualization over another, whereby alternative visualizations would be recommended to the user upon predicting failure. However, for a user who is already deeply engaged and invested in a visualization, a drastic change to an alternative may reset that user’s progress and the task at hand backwards to its initial state. In an attempt to explore adaptations beyond what may be considered as potentially abrupt and intrusive changes, other research efforts have investigated visual modifications to an existing visualization under use, such as highlighting certain visual elements and modifying axes [68], as well as adaptive visualization overlays [67]. In an attempt to facilitate such unintrusive personalization in future adaptive ontology visualization systems, the predictions presented in this paper aim to capture potential user failure before task completion, which in turn will enable real-time adaptations. For example, instead of switching to an entirely different visualization, an adaptive system upon detecting user failure, may apply highlights to sections of the ontology visualization with the goal of directing user attention to relevant ontological hierarchies, entity relationships, and property restrictions to assist with tasks such as ontology development. Likewise, when visualizing evolutionary changes in ontologies, modifications may be made to visual elements such as axes in timelines to potentially offload users’ cognitive efforts to their perceptual systems. Tooltips in ontology visualizations could also provide additional suggestions that direct users to engage with perspective interactive components to reveal additional details. In essence, an adaptive ontology visualization system would extend beyond sole reliance on users’ abilities/initiatives to engage, but rather the system would push/prompt useful visual cues presented within a visualization.
In order to perform such adaptations, a first step is to correctly infer if a user is experiencing issues, e.g., determining their predicted success and failure. Only when it is possible to successfully recognize those moments where a user may possibly benefit from system intervention, can an adaptive visualization provide additional visual cues and assistance to better direct the user’s visual attention in the adaptation. To this end, this paper demonstrates and compares four different approaches to gaze analytics in the prediction of user success and failure, in view of recognizing those moments when an adaptive ontology visualization system could initiate visual adaptation for a specific user. Experimental results generated from a series of trials indicate the feasibility of four potential implementations to predict user success and failure by utilizing eye gaze during user interaction with ontological class visualizations. Most notably, irregular gaze events are found to be most often associated with earliest predictions of user success and failure, while an increase in gaze data is typically found to be associated with higher prediction accuracies over time. In addition, cognitive workload appears to be most indicative of users’ overall success and failure. The contribution of this paper lies in the new knowledge generated on gaze-enabled user predictions, its feasibility to inform the timing of visual adaptation, and the various implementations of predictive gaze analytics leading to differing predictive outcomes in the context of realizing future adaptive visualization systems for ontologies. The results presented in this paper aim to bring us one step closer to achieving adaptive ontology visualization in the near future.

2. Related Work

Visualization along with interactive support for the human user has remained a central theme in semantic web research, and their presence is evident throughout various aspects of semantic technologies traversing from semantic content generation and exploration, to querying and discovery. Early efforts to conceptualize complex semantics date back to first-order logic representation [17]. In related areas such as linked data visualization, an extensive survey [14] on over 70 visualization tools indicates that there is no one best solution, but rather an appropriate solution depending on the specific user goal within the context of a visual exploration. Other related research in knowledge graph visualization has emphasized on improving performance issues surrounding the rendering of large-scale graph visualizations [15], as well as configurable visual explorations for lay users [16]. In ontology visualization, a significant amount of research effort has focused on developing tools and techniques to assist with visual representations of ontologies. One survey [12] classifies 34 ontology visualization techniques into six categories such as indented list, node-link and tree, zoomable, space-filling, focus with context or distortion, and 3D information landscapes, based on the presentation, interaction, functionality, and dimensions utilized in the visualizations. Building on this work, another survey [13] groups 33 ontology visualization tools into largely five categories such as 1.5D, 2D, 3D, and temporal dimensional techniques based on the dimensions, graphical elements, and layout utilized in the visualizations. Examples of recent advances in ontology visualization include efforts to define a visual language for ontologies [18], improving interactions with large-scale ontologies [19, 20], visualizations for novice users without significant knowledge in ontologies [21-23], visualizing changes in ontology evolution [24-27], and ontology visualization recommender systems [28, 29]. One observation is that existing ontology visualization systems typically adopt the one-size-fits-all approach, ignoring individual differences in user needs and preferences. Though some ontology visualization systems such as [30] provide multiple coordinated views (supporting different user preferences) as well as collaborative and social features (supporting asynchronous collaborations with shareable and publishable workspaces), approaches that attempt to adapt ontology visualizations to an individual user remain relatively scarce. This presents an opportunity to add to the existing body of knowledge in this field by investigating potential means to adapt ontology visualizations for an individual user.

Adaptive technologies have long been established in fields such as e-learning and adaptive textbooks [31, 47], recommender systems [32, 48], personalized information retrieval [33, 49], and the adaptive Web [50], to name just a few. Most relevant to ontology visualization, user adaptive systems in the field of InfoViz date back several decades. Early adaptive InfoViz systems such as [34] have investigated personalized visualizations based on the given task or data properties and demonstrated improved user performance in the context of visual information processing. Later research in dynamic adaptive InfoViz systems such as [35] have focused on user expertise and preferences established in prior interactions, to then recommend the most suitable visualizations for
specific users in subsequent interactions, with experimental results demonstrating improved user accuracy and efficiency. Other efforts in real-time adaptive InfoViz systems such as [36] have focused on monitoring users and recommending alternative visualizations based on mouse clicks during an interaction. Since prior interactive data may not be available in every scenario and mouse clicks require direct user input, recent research efforts in adaptive InfoViz have investigated other approaches to realize user adaptation based on indirect user input such as eye gaze [37-39], building on the notion that gaze patterns convey measures of visual attention. The adaptive interface in [37] utilizes eye gaze in the prediction of users’ intentions, whereby a user was asked to drag and drop one visual object to a specified location on the screen, and the adaptation then tailored the content displayed on the screen according to the predicted user goals. The tasks and the predictive goals shown in this paper are distinctively different in comparison. Other research towards adaptive visualization systems [38] has outlined a number of factors in the individual differences that may contribute to effective adaptations, such as personality traits and cognitive abilities, with eye tracking results indicating that users may benefit from different visual presentations and styles in the process of critical information identification and extraction. [39] makes use of deep learning and views eye gaze as sequential data in the prediction of user success in visual search tasks to achieve higher-than-baseline accuracy and proposes a Multivariate Long Short Term Memory Fully Convolutional Network in the use of adaptive systems. In contrast, this paper aims to demonstrate that user predictions do not need to rely on state-of-the-art machine learning algorithms, and that commonly used classification techniques without specialized configuration can also achieve accurate predictions. As such, this work aims to validate that eye gaze can offer an informative source of input in recognizing those users who may be struggling with a given ontology visualization, which will likely inform the design and development of future adaptive ontology visualization systems.

Eye gaze is relatively unique to an individual, and it is inherently unbiased and hyper-personalized much like fingerprints. As such, eye gaze provides an exciting opportunity for adaptive systems in personalized visualization. Its application can be traced back to well-established fields such as cognitive and perceptual psychology as an instrumentation to quantify user attention during information processing [40], reading and searching activities [41]. In InfoViz, eye tracking has been applied to identify pattern variance in visualizations [42] and task types [43], to measure cognitive efforts [44], to predict user interests [45], and to provide visual guidance in comprehension of textual documents [46]. As eye tracking research advances to include standard cameras found on mobile phones, notebooks, and webcams [51-54], as well as infrared eye trackers becoming more affordable with price points below those of smartphones, the landscape of gaze-driven human-computer interaction is rapidly evolving. With emergent and ubiquitous use of eye tracking already evident in gaming, virtual reality, and accessibility research [55-57], it is reasonable to anticipate increased adoptions of gaze-based interactive systems.

In the field of ontology visualization, eye tracking has been utilized as an evaluation tool to assess usability issues of established ontology visualization techniques [58]. The benefit of utilizing eye tracking in the evaluation of a visualization technique is that direct measures such as user visual search and processing behaviors and patterns can be revealed to complement indirect measures such as user success, time on task, and overall satisfaction that implicitly reflect the usability of a visualization. Comparing to eye tracking used as a form of evaluation output, efforts to include eye gaze as a source of input to inform adaptive ontology visualization have been somewhat limited. Prior research [59-61] has investigated gaze and mouse clicks as input sources to predict user performance in both ad hoc and post hoc scenarios, with encouraging results demonstrating the feasibility of gaze-based user predictions that may serve as the basis to determine correct timings of adaptations. The method proposed in [59] requires direct user input such as mouse clicks, which is a condition that may not always be satisfied in every scenario. For instance, when a user is simply gazing at a visualization, it may become challenging to make predictions on user performance. Subsequent research has focused on utilizing eye gaze and not requiring direct user input in predictions in simulated ad hoc [60] and post hoc analytics [61], whereby combinations of the most influential gaze measures on prediction accuracy were reported. For instance, more successful users were shown to have consistently directed their visual attention with greater changes (less dispersed but larger absolute saccadic angles) when searching for relevant visual cues positioned both nearby and further apart in the visualization (more dispersed but shorter median saccadic lengths), generally spent more time on searching for relevant visual cues, and once found, information was quickly
extracted during an interaction (smaller fixation-to-saccade ratios).

Given the nature of eye tracking, streams of gaze data are steadily generated and will continue to grow until the end of an interaction, whereby gaze data can be treated as time-oriented continuous and orderly data, as well as aggregated or discretized batches preceding before and after each distinct visual need of the user. However, there is limited knowledge on how different gaze analysis techniques may influence the outcome of user predictions. To this end, this paper focuses on the comparison of four different methods to process eye gaze in the context of predicting user success and failure. More specifically, this paper emphasizes on viewing eye gaze as either digests and snapshots of data that can be scheduled to evolve an adaptive system, as a basis to detect notable visual events experienced by a user, or as accumulated information sources that reveal all known visual tendencies of a particular user. Based on these different views to eye gaze, this paper presents a series of classification experiments and compares the accuracies of user predictions when utilizing these aforementioned approaches to gaze analytics.

3. Controlled Eye Tracking User Studies

The goal of the controlled eye tracking user studies is to simulate an environment where users must gaze and interact with the ontology visualizations in order to complete a set of tasks. As such, participants were asked to evaluate a set of given mappings between an ontology pair assisted by their visualizations. These user tasks are similar to how automated mapping algorithms are typically evaluated, whereby the automatically generated mappings are compared against a gold standard to determine the precision, recall, and F-measure for an algorithm.

3.1. Experimental Setup

In the experiments, a participant needed to evaluate whether a given mapping was indeed correct in the form of true or false questions, and generate additional mappings if the given set was deemed to be incomplete. Fig. 1 shows two sets of mappings given to the participants to evaluate in the studies (Fig. 1a shows the “Conference domain” and Fig. 1b shows the “Biomedical domain”). The answers from each participant were then compared against the mapping gold standard to determine a correctness success score (i.e., the ratio of correct answers in true or false questions), a completeness success score (i.e., the ratio of correct answers in the newly added mappings), and an overall success score (i.e., the ratio of all correct answers in true or false questions as well as new mappings added against the gold standard). All three types of success scores range between 0 and 1, with 0 being completely unsuccessful and 1 being completely successful.

![Part 1 - Are the Mappings below correct?](image)

**Part 1 - Are the Mappings below correct?**

<table>
<thead>
<tr>
<th>Source Ontology</th>
<th>Target Ontology</th>
<th>True or False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poster</td>
<td>Poster_Paper</td>
<td>True</td>
</tr>
<tr>
<td>Person</td>
<td>Person</td>
<td>False</td>
</tr>
<tr>
<td>Working_event</td>
<td>Scientific_Event</td>
<td>False</td>
</tr>
<tr>
<td>Conference</td>
<td>Conference</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Paper_Author</td>
<td></td>
</tr>
<tr>
<td>Banquet</td>
<td>Conference_Banquet</td>
<td></td>
</tr>
<tr>
<td>Workshop</td>
<td>Workshop_Session</td>
<td></td>
</tr>
<tr>
<td>Topic</td>
<td>Research_Topic</td>
<td></td>
</tr>
<tr>
<td>Contribution</td>
<td>Paper</td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>Conference_Participant</td>
<td></td>
</tr>
<tr>
<td>Chair_PC</td>
<td>OC_Chair</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Organization</td>
<td></td>
</tr>
<tr>
<td>Trip</td>
<td>Conference_Trip</td>
<td></td>
</tr>
<tr>
<td>Member_PC</td>
<td>PC_Member</td>
<td></td>
</tr>
<tr>
<td>Scholar</td>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>Event</td>
<td>Track</td>
<td></td>
</tr>
</tbody>
</table>

**Part 2 - Add Missing Mappings Below (if any):**

<table>
<thead>
<tr>
<th>Source Ontology</th>
<th>Target Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast_form</td>
<td>Yeast</td>
</tr>
<tr>
<td>Genus_HantaVirus</td>
<td>HantaVirus</td>
</tr>
<tr>
<td>Genus_FlawVirus</td>
<td>FlawVirus</td>
</tr>
<tr>
<td>West_Nile_virus</td>
<td>West_Nile_Virus</td>
</tr>
<tr>
<td>Microorganism</td>
<td>Microorganism</td>
</tr>
<tr>
<td>Filamentous_fung</td>
<td>Melt</td>
</tr>
<tr>
<td>Virus</td>
<td>Virus-RNA</td>
</tr>
<tr>
<td>Animal_virus</td>
<td>Animal_Virus</td>
</tr>
<tr>
<td>Bacterial_virus</td>
<td>Virus_Bacterial</td>
</tr>
<tr>
<td>Human_virus</td>
<td>Hepatitis_Virus</td>
</tr>
<tr>
<td>Plant_virus</td>
<td>Virusas_Plant</td>
</tr>
<tr>
<td>Kingdom_Protaryla</td>
<td>Protaryla</td>
</tr>
<tr>
<td>Superkingdom_Bacteria</td>
<td>bacteria</td>
</tr>
<tr>
<td>Bacterium</td>
<td>bacteria</td>
</tr>
<tr>
<td>Hepatitis_O_virus</td>
<td>HDV</td>
</tr>
</tbody>
</table>

**Part 2 - Add Missing Mappings Below (if any):**

<table>
<thead>
<tr>
<th>Source Ontology</th>
<th>Target Ontology</th>
</tr>
</thead>
</table>

![Biomedical Domain](image)

![Conference Domain](image)

Fig. 1. Participants were asked to answer true or false questions for each given mapping (of ontology classes exclusively) and create new mappings if the existing set was deemed to be incomplete.

The ontology pairs and their mapping gold standards were taken from the Ontology Alignment Evaluation Initiative (OAEI), where two different domains including the Conference (Conf) Track and the Biomedical Ontologies (BioMed) Track were used in the controlled studies presented in this paper [1]. The
source and target ontologies in the Conference domain have 74 (38 classes, 13 object properties, and 23 datatype properties) and 110 entities (77 classes and 33 object properties) respectively. The source and target ontologies in the Biomedical domain have 89 and 181 classes (no other types of entities but include 2 and 4 occurrences of multiple inheritances) respectively. The mappings given to the participants were generated based on the gold standards but modified to include incorrect and incomplete mappings to simulate a task environment as discussed above. In each domain, the mapping set presented to the participants always contained the same number of correct, incorrect, and incomplete mappings. More specifically, there are 13 correct mappings, 3 incorrect mappings, and 7 incomplete mappings in a mapping set given to the participant in each domain. Such a setup aims to reduce potential influences of additional variables on user performance such as time on task, where if a participant took longer to complete the evaluation, that was not due to this person needing to work through a greater number of mappings in the given task.

Examples of the indented lists used in the studies can be found in Fig. 2 (for the Conference domain). The node-link diagrams were implemented using the D3 Javascript library [3]. Examples of the node-link diagrams are shown in Fig. 3 (for the Biomedical domain). Finally, Fig. 4 shows the screen layout of the tasks and the accompanying visualizations of a given ontology pair presented to a participant in the study.

![Fig. 2. Examples of the indented list visualization, where classes of an ontology pair in the Conference domain were visualized to the participants to assist with the tasks shown in Figure 5a.](image)

![Fig. 3. Examples of node-link diagrams, where classes of an ontology pair in the Biomedical domain were visualized to the participants to assist with the tasks shown in Figure 5b.](image)

![Fig. 4. Layout of the tasks and visualizations on one computer screen shown to each participant.](image)

To assist with these evaluation tasks, the participants were given either the indented list or the node-link diagram visualization for an ontology pair. These visualization techniques are included in the experiments because they are most frequently used by existing tools as discussed in the related work section. The indented list visualization is achieved by utilizing the Protégé ontology editor [2], where the participants were not allowed to use any other features but solely rely on the visualization of classes.
symbols to illustrate collapsed and expanded nodes in
an ontology. Indentation may be supported by guid-
elines or distinctive styles of bullet points to illustrate
hierarchical levels. Likewise, node-link diagrams can
be arranged into tree structures, cyclic and radial dia-
grams, where collapsed and expanded nodes can be
illustrated by color, style, etc. Recognizing the many
instances in the implementation of a visualization
technique, the specific visualizations used this paper
should be understood as examples of established
techniques when visualizing ontological data. These
visualizations are generalizable in the sense of their
key visual characteristics, whereby indented lists use
indentation to illustrate class hierarchy and a list of
nodes to illustrate the classes within an ontology, and
node-link diagrams use nodes to illustrate classes and
connecting edges to illustrate relationships among
them. These visualizations are not meant to be ex-
haustive, but rather as general examples that are suf-
ficient for the purpose of validating gaze-enabled
predictions in the context of user interaction with
commonly used visualization techniques for ontolog-
ical data models.

To minimize learning effects, each participant in-
teracted with one type of visualization and one do-
main. For example, a participant evaluated mappings
in the Conference domain supported by the indented
lists, and then evaluated another set of mappings in
the Biomedical domain assisted by the node-link dia-
grams. This person did not engage with the same
visualization nor a specific domain twice during that
experimental session. This setup aims to reduce po-
tential bias in user performance (such as time on task
and task success) after becoming familiar with either
given visualization of domain over time. Further-
more, to minimize order effect, the sequence at
which visualizations and their associated domains
presented to the participants were shifted across, so
that overall, participants interacted with visualiza-
tions and domains in various orders. This setup aims
to reduce potential bias in user performance due to a
particular task order, i.e., the same visualiza-
tion/domain always appearing first/second, leading to
unintended impact (e.g., fatigue in the second task)
on the outcome of user performance.

3.2. Data Collection

In addition to the three types of user success dis-
cussed in the previous section, user gaze was collect-
ed for each participant throughout an interactive ses-
sion. To achieve this, participants were seated in
front of a computer monitor (21.3” screen with
1600×1200 resolution) that has a built-in eye tracker
(Trusbi 2150). No additional head mounts or sensory
hardware were used. This physical configuration is
representative of common scenarios of human-
computer interaction and simulates realistic real-
world environments of interactions with ontology
visualizations. Each participant calibrated with the
eye tracker before each study session to ensure max-
imum data validity. All studies were conducted in the
same computer lab with constant indoor lighting.
Participants were seated on non-wheeled and non-
swivel chair, so that they remained in the same dis-
tance to eye tracking sensor after successful cali-
bration. Participants were informed that there were cor-
correct and wrong answers to the given questions and
instructed to complete them as quickly as possible
but without a preset time limit.

A total of 36 participants took part in the studies,
which included undergraduate and graduate students
majoring in Psychology, Computer Science, Biomed-
ical Engineering, Mechanical Engineering, and Elec-
trical Engineering. Before each session, all partici-
pants were given the same tutorial (based on the Piz-
za ontology [4]) on ontologies and briefed on how to
interact with an indented list or a node-link diagram.
Participants were instructed to raise questions prior to
the start of each task session, and once they had be-
gun, they were not allowed to interact with the re-
searcher running the experiment any longer. This
setup aims to protect the validity of the user data col-
lected such as time on task. To minimize noise in the
gaze data collected for each participant, invalid data
entries were discarded, including those reported by
the eye tracker, entries missing pupil coordinates, as
well as those associated with pupil dilations that were
outside of the possible ranges. The American Acad-
emy of Ophthalmology reports that normal pupillary
size range between 2-8mm in size. More specifically,
the normal pupil size in adults varies from 2.4mm in
diameter in bright light to 4-8mm in the dark [5].
Finally, due to eyeglasses and contact lenses, some
participants generated corrupted data that were dis-
carded entirely. Overall, after the data cleaning pro-
cess, a total of 33 participants’ data were used in the
predictive analytics presented in this paper.

3.3. Gaze Metrics & Gaze Feature Sets

An overview of a set of basic descriptive gaze
metrics is presented in Fig. 5 and Table 1. Fixations
refer to periodic moments throughout an interaction when a person’s gaze is relatively stationary as this person fixates visual attention on a point of interest. Each fixation can be associated with a duration (measured in milliseconds), which may indicate a user’s need to process visual information during that time. Typically, shorter fixations on average likely indicate faster information extraction and processing by the user [6]. As an interaction concludes, it is possible to determine the total number of fixations occurred for a specific user. This count of fixations (measured as positive integers) may be indicative of how users spent their visual attention during an interactive session, whereby fewer total fixations may suggest more effective visual cues in the visualization [6].

As users direct their gaze between various fixations, the quick eye movement from one fixation to the next is referred to as saccades. Similar to fixations, saccades can be associated with durations and counts (measured in milliseconds and positive integers respectively). In addition, the distance between successive fixations is referred to as saccadic length (measured in pixels). Saccadic lengths are typically indicative of the placement of relevant visual components, whereby longer saccadic lengths may suggest fewer unnecessary interim fixations [6]. Upon learning the total fixation and saccade durations during an entire interaction, it is therefore possible to determine the saccade-to-fixation ratio, known as the search-to-process ratio. This ratio aims to capture an overall impression of how users spent their visual attention during an interaction. A ratio at 1 indicates equal time spent on searching and processing information, a value less than 1 indicates more time spent on processing information, and a value greater than 1 indicates more time spent on searching for information during an interaction.

As users scan for relevant visual cues and change directions in their visual searches, it is possible to measure such changes as relative and absolute saccadic angles (in degrees). Relative saccadic angles indicate the degrees of change in successive fixations (an example is shown in Fig. 5). Relative saccadic angles typically indicate veering from a set course [7], where angles greater than 90° are typically considered as backtracks indicating potential changes in user goals or misaligned user expectations and the visualization itself [6]. Absolute saccadic angles refer to the degrees of change with respect to the global horizontal axis (an example is shown in Fig. 5). Absolute angles typically indicate spatial tendencies relative to the element layout on a page [7], such as user tendencies in their searches of sinistrodextral entity labels shown in the ontology visualizations used in this paper.

The complete scanning behavior of a user from the start to the end of an interaction yielding a collection of fixations and saccades is referred to as a scanpath. Scanpaths are associated with durations (i.e., the sum of all fixation and saccade durations, measured in milliseconds) and lengths (i.e., the sum of all saccadic lengths, measured in pixels). The smallest bounding area containing all fixations captured dur-

![Fig. 5. Descriptive gaze metrics, illustrating fixations, saccadic lengths, relative and absolute angles.](image-url)

<table>
<thead>
<tr>
<th>Table 1. Descriptive Gaze Measures Collected Per Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indication</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Indicative of Information</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Indicative of Information</strong></td>
</tr>
<tr>
<td><strong>Processing Activities</strong></td>
</tr>
<tr>
<td><strong>Indicative of Cognitive Workload</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
ing an entire interaction is referred to as the convex hull area (measured in px²). These measures are often interpreted with respect to one another. For instance, a smaller convex hull area coupled with shorter scanpath lengths may suggest a more effective visualization [6].

Pupil dilation has been used as an indicator of cognitive workload [8] with increased pupil sizes signifying more cognitive efforts demanded [9, 10] in a given task. A user’s pupil sizes captured during an interaction can be compared to this person’s baseline value in order to determine pupil dilation. In this paper, a user’s baseline pupil sizes were collected prior to the start of a given task (during calibration).

The aforementioned gaze metrics can be further grouped into four categories as follows, and are referred to as gaze feature sets in this paper henceforward:

- gaze measures that indicate information search activities, including count of fixations and saccades, saccadic length and duration, convex hull, and scanpath length;
- gaze measures that indicate information processing activities, including fixation duration and search-to-process ratio;
- gaze measures that indicate cognitive workload, including pupil dilation, relative and absolute saccadic angles; and
- the complete set of gaze measures that include all metrics as defined above.

In the experiments presented in this paper, these four gaze feature sets are combined with four different approaches to predictive gaze analytics (discussed next).

4. Predictive Gaze Analytics

The goal of the experiments presented in this paper is to compare four different approaches in the context of predictive gaze analytics that focus on learning from digests of user gaze (section 4.1), snapshots of the most recent user gaze (section 4.2), gaze captured during significant events (section 4.3), as well as all known gaze to date (section 4.4). There are a number of possible approaches to analyze gaze data, and these four implementations presented in this paper are not designed to be exhaustive, but rather as examples of different approaches that emphasize on various types of gaze state when making predictions on user success and failure. In each approach presented below, the gaze feature sets discussed in section 3.3. (i.e., complete gaze features, feature sets that indicate search activities, processing activities, and cognitive workload) were generated for each user.

4.1. Scheduled Gaze Digests

One approach to generate continuous gaze digests during a user’s interaction with ontology visualizations is to analyze gaze data in a series of non-overlapping, fixed-size windows at scheduled contiguous time intervals. An example of scheduled gaze digests is shown in Fig. 6, where we make initial predictions based on user gaze found in a given time zone (i.e., window 1 in Fig. 6), then move on to the next bordering time zone to make subsequent predictions (i.e., window 2 in Fig. 6), and tumble forward until we reach the end of an interaction. The window size, or time zone, can be defined at any interval as appropriate depending on the given scenario. In this paper, scheduled gaze digests are empirically mapped to two-minute tumbling windows in the experiments. This approach to predictive gaze analytics takes a digest view on the users’ gaze data and focuses on providing scheduled non-overlapping updates in the process of generating predictions.

![Fig. 6. Taking a scheduled digest view of the gaze data using a tumbling window that is non-overlapping and fixed in size](image)

4.2. Recent Gaze Snapshots

It can be argued that taking a digest view on user gaze ignores potential relationships and dependencies between chunks of data since this approach disregards previously reported gaze behaviors, which may or may not reflect the most current state of user attention as it evolves during an interaction. As such, another approach to predictive gaze analytics is to take a snapshot view on the gaze data that aims to capture the most recent state of user attention by utilizing an overlapping hopping window such as the example illustrated in Fig. 7, whereby scheduled overlapping windows are utilized at a given interval. More specifically, a hopping window analyzes gaze in a defined window size to generate predictions, then moves
forward to the next scheduled hop relative to the previous one. In this paper, we used a two-minute hopping window with a one-minute hop size, i.e., every minute, we analyzed gaze over the last two minutes to make predictions on user success and failure. The goal of this approach is to capture a series of continuous snapshots that reflect the most recent gaze state of the user.

![Fig. 7](image)

**Fig. 7.** Taking the most recent snapshot view of the gaze data using a hopping window that is overlapping and fixed in size

### 4.3. Event-Based Gaze Patterns

During interactions with ontology visualizations, users may experience defining moments that are impactful enough to affect their performance to a notable degree. These defining moments may translate to notable gaze behaviors resembling phases of significant events, which are likely to be indicative of a user’s performance. Within this context, notable events can be defined as anything unusual relative to what is already known of a user’s gaze, such as a backtrack, a change in gaze direction, or pupil dilation. There are a number of potential ways to implement notable events; this paper, a notable event is defined as a longer-than-usual fixation. More specifically, a user’s average fixation duration after the first two minutes of interaction is used as a baseline, whereby all future fixation durations would be compared against, and deemed as notable once found longer than this baseline. We used a non-overlapping, non-fixed-size session window (as shown in Fig. 8) to achieve this approach to predictive gaze analytics. More specifically, a session window begins when the first event is found; it then keeps searching for the next event within a specified time period. If nothing is found, it would close at a specified time out (e.g., window 1 in Fig. 8); if another event is found, the session window would extend the search within another timeout period and repeat this process (e.g., window 3 in Fig. 8) until a specified maximum duration (e.g., window 2 in Fig. 8). In this paper, the session window is mapped to a two-minute timeout and a five-minute maximum duration. The goal of this approach to gaze analytics is to emphasize on potentially more informative chunks of gaze data in the predictions of user success and failure, where the weight is placed on notable events rather than taking a scheduled view on gaze data stream.

![Fig. 8](image)

**Fig. 8.** Taking an event-based view of the gaze data using a session window that is non-overlapping and non-fixed in size

### 4.4. Cumulative Gaze Behavior

As a user interacts with an ontology visualization to complete a given task, this person’s gaze data continues to grow until the task is concluded and the interaction terminates. This provides an opportunity to analyze all gaze collected for a user in the predictive analytics. To this end, we implement an expanding window that takes a cumulative view and considers all that is known for a person’s gaze when making predictions (as shown in Fig. 9). An initial set of gaze data collected from a user is analyzed (e.g., window 1 in Fig. 9), we then expand the predictive analytics to include new gaze data generated for that person at the next specified time interval (e.g., window 2 and 3 in Fig. 9). In this paper, the expanding window has a scheduled two-minute interval, whereby new data is added to previously known gaze data to inform the predictive analytical process. The goal of this approach is to emphasize on all gaze that is known to the system when generating predictions on user success and failure.

![Fig. 9](image)

**Fig. 9.** Taking a cumulative view of the gaze data using an expanding window that is overlapping and non-fixed in size
5. Results

Based on a median split of user success, the classifications shown in this paper aim to predict whether a user’s performance belongs to the high (i.e., above the median split) or low (i.e., below the median split) group for that measure. Off-the-shelf classifiers taken from the Waikato Environment for Knowledge Analysis (WEKA) machine learning toolkit [11] were used in the experiments shown in this paper. A range of example classifiers representative of established classification models were applied using stratified 10-fold cross validation for model evaluation and Bonferroni-corrected t-tests for statistical testing. These example classifiers included tree-based, e.g., reduced error pruning tree (REPTree); regression-based, e.g., classification via regression (REG); probabilistic, e.g., Bayesian Network (BayesNet); and neural network, e.g., Multilayer Perceptron (MLP). The results generated from these models were compared to those of a commonly used baseline, namely zero rule (ZeroR) that predicts the most frequently occurring category. The purpose of having a baseline classifier is to provide a necessary benchmark to compare various approaches to gaze analytics (discussed in section 4). It is important to recognize that the goal of the experiments is not identify one best machine learning algorithm, but to demonstrate that a range of standard classifiers can perform well without specialized configurations or expert knowledge in machine learning, which highlights the feasibility of incorporating such predictive analytics in future adaptive ontology visualization systems.

Table 2 presents the accuracies of user predictions (on their success and failure to determine if mappings are correct, complete, as well as their overall accomplishment in assessing these aspects of a given mapping set) generated by various example classifiers (using BayesNet, MLP, REG, REPTree classification models), organized by the type of gaze feature sets used (i.e., complete gaze features, those that indicate searching or processing activities, and cognitive workload). Each chart presents accuracies achieved throughout an interactive session (i.e., the amount of time users spent interacting with a given ontology visualization) based on the type of approach used to analyze user gaze (i.e., either taking a digest, snapshot, event-based, or cumulative view to predictive gaze analytics). In addition, for each chart shown in Table 2, its corresponding report on the statistical significance of the accuracies compared to the baseline can be found in Table 3. It is important to note that not all users completed the given tasks using the same amount of time, though the charts show results up to 20 minutes in Table 2 and 3. This is because at least half of the participants took 20 minutes or longer to complete the given tasks. As such, the results generated up to this point in time demonstrate predictions for at least half of the user group, as opposed to potentially non-presentative results based on a minority group of individuals who remain unfinished in their tasks as more time passes by.

5.1. Predicting Users’ Correctness Success

When predicting users’ success and failure at assessing whether given mappings are correct, the baseline classifier generated predictions with accuracies in the range of 54.76–62.86%, where a peak value was achieved after 14 minutes using the event-based analytics coupled with the complete gaze dataset. As shown in Table 2 and 3 (second columns), statistically significant higher accuracies in the range of 68.14–81.15% were generated across all couplings of analytical approaches and gaze features, where the highest accuracy was found using the digest approach coupled with the complete gaze dataset. Most notably, higher-than-baseline predictions were generated as early as after 2 minutes of interaction, using the event-based analytical approach coupled with either the complete gaze features or those indicating cognitive workload. Also, the accuracies of the predictions increased with longer interactions, whereby irrespective of which analytical approach or gaze feature sets used, higher accuracies as well as more numbers of statistically significant results were found after approximately 16 minutes of interaction.

5.2. Predicting Users’ Completeness Success

When predicting users’ success and failure at evaluating the completeness of a given mapping set, the baseline achieved predictions with 44.00–63.52% accuracies (highest value was generated after 20 minutes of interaction using the event-based analytical approach coupled with complete gaze features). As shown in Table 2 and 3 (third columns), significantly higher accuracies were generated throughout the interactive process at every interval across all couplings of the analytical approach and the gaze feature sets. Different from predictions on users’ correctness success and failure, more numbers of statistically significant predictions were found at the beginning of an interaction (e.g., after 2 minutes) as
opposed to later stages. The highest accuracy found was 78.33% after 18 minutes of interaction using the snapshot approach coupled with the complete gaze features.

Table 2. Results of Prediction Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Correctness</th>
<th>Completeness</th>
<th>Overall Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Gaze Features</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td>Gaze Features Indicating Searching Activities</td>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td>Gaze Features Indicating Processing Activities</td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
<td><img src="image9" alt="Graph" /></td>
</tr>
<tr>
<td>Gaze Features Indicating Cognitive Workload</td>
<td><img src="image10" alt="Graph" /></td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
</tbody>
</table>
5.3. Predicting Users’ Overall Success

When predicting users’ overall success and failure at the given tasks, the baseline accuracies were found to be in the range of 50.71–62.00%, where the highest value was generated after 16 minutes of interaction using the digest approach coupled with gaze features indicating processing activities. As shown in Table 2 and 3 (fourth columns), higher accuracies in the ranges of 69.74–78.58% were found in various analytical approaches using all types of gaze feature sets. The peak value was found at a number of settings and always after 18 minutes of interaction, including combining the cumulative approach with complete gaze features or those indicating search activities, as well as coupling gaze features that either indicate processing activities or cognitive workload with any analytical approach except event based. Notably, the earliest and consistently higher accuracies were found using the digest analytical approach coupled with gaze features that indicate cognitive workload.

Table 3. Significance of Predictions (p<0.05)

<table>
<thead>
<tr>
<th></th>
<th>Correctness Success</th>
<th>Completeness Success</th>
<th>Overall Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Gaze Features</td>
<td><img src="chart1" alt="Graph" /></td>
<td><img src="chart2" alt="Graph" /></td>
<td><img src="chart3" alt="Graph" /></td>
</tr>
<tr>
<td>Gaze Features Indicating Searching Activities</td>
<td><img src="chart4" alt="Graph" /></td>
<td><img src="chart5" alt="Graph" /></td>
<td><img src="chart6" alt="Graph" /></td>
</tr>
<tr>
<td>Gaze Features Indicating Processing Activities</td>
<td><img src="chart7" alt="Graph" /></td>
<td><img src="chart8" alt="Graph" /></td>
<td><img src="chart9" alt="Graph" /></td>
</tr>
<tr>
<td>Gaze Features Indicating Cognitive Workload</td>
<td><img src="chart10" alt="Graph" /></td>
<td><img src="chart11" alt="Graph" /></td>
<td><img src="chart12" alt="Graph" /></td>
</tr>
</tbody>
</table>

![Pie Chart](chart13)
6. Findings and Discussions

The results showed mixed outcomes depending on a number of variables including the type of gaze features, classifiers, and gaze analytical approach used. There was no one best approach to predictive gaze analytics that was universally superior to all others across various scenarios. However, there are a number of notable observations that may be informative to the adaptive process of interactive ontology visualizations, as discussed below. A summary of earliest predictions and peak accuracies by analytical approaches and gaze feature sets are presented in Table 4 and 5 respectively.

| Table 4. Accuracies (%) by Earliest Predictions (Mins) (*Statistically Significant p<0.05) |
|---------------------------------|-----------------|-----------------|-----------------|
|                                 | Digest          | Snapshot        | Event           | Cumulation      |
| Complete Gaze Features          |                 |                 |                 |                 |
| Accuracy                        | 68.14*          | 72.00*          | 62.76           |
| Time Interval                   | 2               | 2               | 4               |
| Gaze Analytics                  |                 |                 |                 |                 |
| Gaze Features Indicating        |                 |                 |                 |                 |
| Searching Activities            |                 |                 |                 |                 |
| Accuracy                        | 63.07           | 72.00*          | 56.81           |
| Time Interval                   | 2               | 2               | 4               |
| Gaze Analytics                  |                 |                 |                 |                 |
| Gaze Features Indicating        |                 |                 |                 |                 |
| Processing Activities           |                 |                 |                 |                 |
| Accuracy                        | 63.07           | 71.83*          | 51.95           |
| Time Interval                   | 2               | 2               | 2               |
| Gaze Analytics                  |                 |                 | Digest/Snapshot/Event/Cumulation |
| Gaze Features Indicating        |                 |                 |                 |                 |
| Cognitive Workload              |                 |                 |                 |                 |
| Accuracy                        | 67.74*          | 72.00*          | 67.71*          |
| Time Interval                   | 2               | 2               | 18              |
| Gaze Analytics                  |                 |                 | Digest          |

| Table 5. Peak Accuracies (%) by Time Intervals (Mins) (*Statistically Significant p<0.05) |
|---------------------------------|-----------------|-----------------|-----------------|
|                                 | Digest          | Snapshot        | Event           | Cumulation      |
| Complete Gaze Features          |                 |                 |                 |                 |
| Accuracy                        | 81.15*          | 78.33*          | 78.58*          |
| Time Interval                   | 16              | 18              | 18              |
| Gaze Analytics                  |                 |                 | Digest/Snapshot/Cumulation |
| Gaze Features Indicating        |                 |                 |                 |                 |
| Searching Activities            |                 |                 |                 |                 |
| Accuracy                        | 80.33*          | 77.42*          | 78.58*          |
| Time Interval                   | 18              | 18              | 18              |
| Gaze Analytics                  |                 |                 | Digest/Snapshot/Cumulation |
| Gaze Features Indicating        |                 |                 |                 |                 |
| Processing Activities           |                 |                 |                 |                 |
| Accuracy                        | 79.58*          | 73.58*          | 78.58*          |
| Time Interval                   | 18              | 18              | 18              |
| Gaze Analytics                  |                 |                 | Digest/Snapshot/Cumulation |
| Gaze Features Indicating        |                 |                 |                 |                 |
| Cognitive Workload              |                 |                 |                 |                 |
| Accuracy                        | 80.33*          | 73.58*          | 78.58*          |
| Time Interval                   | 18              | 18              | 18              |
| Gaze Analytics                  |                 |                 | Digest/Snapshot/Cumulation |

Predictions based on notable gaze events are often generated earlier than that of others – the event-based approach to gaze analytics seems to yield predictions that are often found at the beginning of an interaction (i.e., 2–4 minutes), as shown in Table 3. This is observed in a number of cases and most notably, when predicting users’ correctness success (using either the complete gaze features or those that indicate cognitive workload) and overall success (using either the complete gaze feature sets or those that indicate searching activities). This finding may be speculated in how notable gaze events are defined in this paper, i.e., when a user’s fixation duration exceeds what can be considered as typical for that person (based on the average fixation duration known after 2 minutes of interaction). It is reasonable to anticipate that users would need to become acquainted with a given visualization at the beginning of an interaction with most attention spent on looking around a visual environment. As such, this process is likely to lead to gaze features that may be perceived as exhibitions of search behaviors or resemblances of cognitive demand, while being relatively comparable to the complete gaze dataset captured for a person at the beginning of an interaction. When coupled with event-based gaze analytics, these combinations would therefore likely lead to earlier predictions.

Cumulated gaze analytics are often associated with higher accuracies during predictions at later interactive stages – it appears that the cumulative approach to gaze analytics often yields predictions with higher accuracies towards the later stages of interactions (i.e., 16–18 minutes) compared to other
analytical approaches, as shown in Table 4. Though these results may not be the earliest indicators of user success and failure (unlike the event-based approach discussed above), this approach is found to be associated with the greatest number of peak accuracies achieved across all types of success predictions. It is consistently overserved in the case of predicting users’ overall success in particular, that irrespective of the gaze feature sets used, the cumulated approach to gaze analytics has generated the highest accuracies compared to that of other approaches. These findings may be speculated in the nature of this analytical approach, whereby the growing gaze features (regardless of whether they indicate certain types of activists or not) are likely to contain richer data to inform the predictive process as more time passes in an interaction. The results shown in this paper should be understood within the context of the specific experimental conditions, as the user predictions are likely to be task and visualization dependent. As such, cumulated gaze may not always lead to increased accuracy as fewer discerning gaze patterns may also develop over time. Furthermore, some scenarios may entail that it is not rational (e.g., if a task only requires a few seconds or minutes to complete) to analyze gaze in this fashion, as later predictions may provide little benefit for timely intervention. As a result, cumulated gaze analytics may be most suitable for post doc adaptations, where a visualization system adapts visual cues displayed to the current user based on prior interactions and predictions learned from past users.

Significantly better results when predicting completeness success – higher accuracies as well as the greatest number of statistically significant results were found across all gaze analytical approaches and feature sets when predicting users’ completeness success or failure (compared to correctness or overall success). Irrespective of which approach to gaze analytics was applied, a large number of predictions were generated early and remained relatively high throughout the interaction. Different from assessing correctness, the accuracies of the predictions on completeness success did not increase much with time, or in other words, longer interaction. In fact, it is almost predicted immediately whether a person is likely to succeed or not. This observation may be a result of the nature of the user task. In order to assess whether a given mapping set is complete or not, a user must gain an overall understanding of the visual semantics at hand, before creating new mappings based on that knowledge. It is likely that regardless of how gaze data was analyzed, each data chunk (be it a digest, snapshot, event, or cumulated set of gaze) presented similar patterns that led to relatively constant predictions. In other words, it may be said that gaze analytics may be more informative on predicting one’s accomplishment at creating new mappings than assessing existing mappings (i.e., easier to predict completeness than correctness).

Cognitive workload appears to be an indicator of overall success or failure – compared to other gaze feature sets, those that indicate cognitive workload seem to be most indicative of overall success, whereby early predictions (and at every interval throughout the interaction) as well as significantly higher accuracies were found when coupled with the digest analytical approach. This finding suggests that a person’s cognitive workload may be the most appropriate indicator on how well this person is performing a given task. Without breaking down the task into different aspects (correctness vs. completeness as discussed above), cognitive workload may be perceived as the most applicable measure that accounts for all that is required to accomplish a given task. As such, non-overlapping scheduled digests of gaze features indicting cognitive workload are likely to be more informative when predicting one’s overall success. It may be necessary to note that this finding may not be evident in every scenario, as indicators of cognitive workload such as pupil dilation can be subjective to environmental factors (e.g., change in lighting).

7. Conclusions and Future Work

This paper presents a comparative study on four different approaches to analyze gaze in view of facilitating future adaptive ontology visualization systems, whereby it is envisioned that accurate predictions on users’ success and failure may be used to inform adaptive systems to potentially tailor visual cues for a specific user. The overall goal of visual adaptations is to enhance the user experience and improve user performance. This study focuses on utilizing both overlapping (i.e., snapshots of user gaze and cumulated gaze over time) and non-overlapping (i.e., scheduled gaze digests and notable gaze events) gaze analytics to generate predictions using off-the-shelf classification models without the need of specialized configurations. The results show that accurate predictions can be generated despite a presence of mixed user backgrounds and expertise. These findings suggest that all four approaches to gaze analytics may be helpful in inferring user predictions, though some
approaches may be more appropriate than others in certain scenarios (e.g., if it is desirable to predict potential outcomes as early as possible such as in a time-sensitive task environment, or if it is important to be as confident as possible in the predictions such as in a specificity-critical environment) as discussed below, along with several limitations as well as the implications of potential future research directions.

In settings where it is essential to extract early indications on how a user is doing with a given visualization, detecting notable events in gaze analytics may be a valuable contributor to rapid predictions of user success and failure. This paper has explored one instance of notable gaze events focusing on changes to a person’s fixation duration (i.e. exhibiting an unusually long fixation duration), though there are many other possible implementations, such as frequencies of gaze backtracks (how often visual cues are scanned and rescanned as users make sense of them), time to first fixation in a given area of interest (how visual attention may be prioritized in a pre-defined space), and sequences of fixation points (in environments where tasks can only be completed with successions of orderly fixations) to name just a few. It is not yet known if sudden changes in other types of gaze events may lead to similar or even better results. Future research could investigate if additional gaze events would be more effective at producing even more accurate and earlier predictions.

In scenarios where higher accuracies are desired, it may be necessary to analyze cumulated gaze after an extended amount of time and interaction. However, such a condition may not be possible in every situation, and particularly when the tasks at hand will only take a few minutes. This creates an opportunity to study the tradeoff (if any) between accuracy sensitivity versus specificity. In addition, the implementations of cumulative gaze analytics can explore beyond regular timed intervals, such as at each point after the user has directed attention to a predefined visual component that is critical to task success. Future research could investigate contributing factors to effective use of cumulatively analyzing all gaze that is known about a user at a given point in time.

In the case of predictions by gaze digests and snapshots, while one takes a non-overlapping strategy and the other takes an overlapping view to gaze analytics, they have one characteristic in common, i.e., gaze is essentially treated as isolated chunks of data independent from what may already be known about a user’s gaze behavior up until a certain moment in time. It is possible that a scheduled timed interval may have broken up the context from which gaze data was generated within. In the experiments presented in this paper, digests and snapshots have not produced exceptional results or notable patterns when compared to other approaches. However, it is not yet known whether other implementations may yield better results if differing techniques were applied when defining the window size and frequency. Future research could potentially focus on identifying the task environments and visual conditions that must be satisfied to optimize utilizations of gaze digests and snapshots.

The timing of predictions as well as their accuracies appear to vary depending on the type of tasks (i.e., correctness, completeness, or overall success). Take the example where the complete gaze features are used, the event-based analytical approach has generated the earliest statistically significant predictions, but the peak accuracy was achieved based on gaze digests when predicting correctness success, while similar results were found using gaze snapshots and the cumulative approach when predicting completeness as well as overall success. This result may be grounded in the differing nature of the tasks, which likely may have led to distinct gaze behaviors. In the case of correctness success, in order to determine if an existing mapping is true or false, a user must first visually locate the given entity pair before making a judgement on whether they should be mapped. In the case of completeness success, a thorough understanding of the entire sets of ontological entities is required to notice anything missing and subsequently creating new mappings. It may be expected that the former is largely driven by rapid discoveries of visual information, while the latter relies on comprehensions of all visual cues presented. When both are accounted in the overall success, it may be reasonable to anticipate mixed outcomes. Future research could potentially investigate whether visual needs are indeed distinctively different in these types of tasks, and if it is possible to correlate visual needs with predictive analytics.

The classification models used in the experiments are established examples of basic learning techniques. While they are sufficient for the purpose of comparing the four analytical approaches, they do not explore specialized configurations or sophisticated learning techniques, which are outside the scope of this paper. Future research could investigate improving accuracies of the learning models in the context of predictive gaze analytics. Furthermore, future studies could also include additional variables to study the impact they may each have on predictive outcomes, such as visualization types (if a particular
gaze analytical approach is most appropriate for a specific types of ontology visualization), user tasks (additional ontological tasks commonly assisted by ontology visualizations such as establishing hierarchical, instance, and multiple inheriting relationships, entity creation, ontology evolution, etc.), and user groups (with distinct participant backgrounds such as domain expertise, visual literacy, cognitive style, and how each attribute many effect on the predictive outcome).

References

[1] Ontology Alignment Evaluation Initiative, tracks and gold standards can be found at http://oaei.ontologymatching.org
[26] C. Ochs, J. Geller, M.A. Musen, Y. Perl, Real Time Summarization and Visualization of Ontology Change in Protégé. CEUR-WS 1947, 75-86. VOILA@ISWC 2017
[27] R. Permnich, M. Serbak, D. dell’Aglio, A. Bernstein, Chimp: Visualizing Ontology Changes and their Impact in Protégé, CEUR-WS 2778, 47-60. VOILA@ISWC 2020
[29] O. Sváb-Zamazal, M. Duidás, V. Svátek, Augmenting the Ontology Visualization Tool Recommender Input Pre-Filling and Integration with the OOSP Ontological Benchmark Builder. SEMANTiCS 2016


