

---

## Supporting user-centred ontology visualisation: predictive analytics using eye gaze to enhance human-ontology interaction

---

Bo Fu\*

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

Email: Bo.Fu@csulb.edu

\*Corresponding author

Ben Steichen

Computer Science,  
California State Polytechnic University Pomona,  
3801 W Temple Ave, Pomona, CA 91768, USA  
Email: bsteichen@cpp.edu

**Abstract:** Visualisation is an important aspect to support human-ontology interaction, as visual cues amplify cognition and offload cognitive efforts to the human perceptual system. While significant research efforts have focused on visualisation layouts, adapting to the individual user has been largely overlooked in typical ontology visualisation systems. This provides an opportunity to potentially seek more personalised support in ontology visualisation. As such, this paper utilises a tumbling window analytical technique and demonstrates accurate predictions of a user's likelihood to succeed in a given task based on this person's latest gaze data during an interactive session. We show several trial scenarios where statistically significant accuracies are achieved for two commonly used ontology visualisations in the presence of mixed user backgrounds and task domains. In addition, depending on the gaze features that emphasise a user's search or processing activities, or cognitive workload, trial results show earlier predictions as well as higher accuracies can be achieved in some cases. Furthermore, an investigation of influential gaze features reveals a combination of gaze traits is often associated with higher user success. These findings motivate and highlight potentially ample opportunities to adapt to the individual user throughout various interactive stages in the realisation of adaptive ontology visualisation.

**Keywords:** semantic web; semantic data interaction; predictive analytics; adaptive ontology visualisation; eye tracking; applied machine learning.

**Reference** to this paper should be made as follows: Fu, B. and Steichen, B. (2022) 'Supporting user-centred ontology visualisation: predictive analytics using eye gaze to enhance human-ontology interaction', *Int. J. Intelligent Information and Database Systems*, Vol. 15, No. 1, pp.28–56.

**Biographical notes:** Bo Fu is an Associate Professor in the Department of Computer Engineering and Computer Science at California State University Long Beach (CSULB), USA. She received her BSc (Joint Honors), MSc and PhD in Computer Science all from Trinity College Dublin, Ireland. Prior to joining CSULB, she was a Research Scientist at the National Center for Biomedical Ontology, and a Postdoctoral Fellow at the University of Victoria, Canada. Her research intersects web semantics and human data interaction. She is interested in accelerating research and innovation in semantic-oriented intelligent systems and interactive visual analytics that facilitate human decision-making.

Ben Steichen is an Assistant Professor in the Computer Science Department at California State Polytechnic University, Pomona (CPP), USA, and his research focuses on human-centred computing, user modelling, and personalised information access. Before joining CPP, he was an Assistant Professor at Santa Clara University, USA. Prior to this, he was a Postdoctoral Research Fellow at the University of British Columbia, Canada. He received his PhD and MSc degrees in Computer Science from Trinity College Dublin, Ireland, and a First Class Honours BSc degree in Computing Sciences from the University of East Anglia, UK.

This paper is a revised and expanded version of a paper entitled ‘Tumbling to succeed: a predictive analysis of user success in interactive ontology visualization’ presented at the 10th International Conference on Web Intelligence, Mining and Semantics, Biarritz, France, 30 June to 3 July 2020.

---

## 1 Introduction

Recent efforts in the field of ontology visualisation have largely focused on developing various graphical tools and producing technical advances in interaction design, such as improving visual navigational support (Motta et al., 2011), creating visualisations for users without semantic technology background (Sabol et al., 2014), and generating recommendations of visual renditions (Theilmann et al., 2015). Although individual users’ visual preferences, information needs, and cognitive abilities are recognised in adaptive and personalised technologies applied to others fields such as information retrieval (Steichen et al., 2012), e-learning (Chen, 2008), and recommender systems (Cho et al., 2002), the adoption of adaptive and personalised visualisation systems remain scarce in ontology visualisation.

Motivated by prior research in adaptive information visualisation systems that has generated promising results and demonstrated valuable benefits of tailoring to individual users based on a range of adaptation methods such as the given task at hand (Casner, 1991; Mackinlay, 1986), user expertise (Grawemeyer, 2006), and real time user observation (Gotz and Wen, 2009), this research aims to facilitate adaptive ontology visualisation by means of identifying opportunities where personalisation may be achieved during human-ontology interaction, i.e. before the user completes a given task. To this end, we utilise a tumbling window technique when analysing a user’s eye gaze data captured during an interaction, with the overall goal of inferring this person’s likely success with a given task in real time. These predictions can then be used to inform the underlying ontology visualisation system, which could subsequently make appropriate

changes to the visual cues presented to the user. These personalised changes may potentially increase a user's success rate, reduce time needed to complete the given task, and improve the overall user experience with the ontology visualisation.

In this paper, we present predictive modelling of users' gaze data using established machine learning techniques and demonstrate several predictions with higher accuracies against those of a baseline classifier. In particular, we present results showing accurate predictions of users' success throughout multiple stages of an interaction and in some cases, as early as a few minutes into an interactive session with ontology visualisations. Extending findings reported in Fu et al. (2020), we investigate several configurations of a tumbling window technique coupled with gaze datasets featuring user activities such as search, processing, and cognitive workload, whereby further improvements were found in user success predictions such as earlier and higher statistically significant accuracies. Finally, we conclude with lessons learned based on additional insights gained from investigating the most influential gaze features. These results demonstrate that it is not only possible to infer a user's likely success with a given task at early stages of an interaction, but also feasible to generate multiple accurate user predictions throughout various stages of a human-ontology interaction. These findings provide further evidence that personalised ontology visualisation may be achieved if inferred user success could be accounted by the underlying visualisation system in the rendering of visual displays.

## 2 Related work

Ontology visualisation plays an important role in human-ontology interaction, and is a thriving research area with continuous development of numerous tools and techniques over the years. Examples of recent research emphasis include providing visual summaries to support navigation and enhance comprehension of large-scale ontologies (Motta et al., 2011; Santana-Pérez, 2018), creating visualisations for casual users without significant backgrounds in semantic technologies (Sabol et al., 2014; Lohmann et al., 2015; Graziosi et al., 2017), visualising ontological changes during editing and evolution (Burch and Lohmann, 2015; Lambrix et al., 2016; Ochs et al., 2017), enriching query languages with visual representations (Zainab et al., 2015; Haag et al., 2015; Bartolomeo et al., 2018; Vargas et al., 2019), as well as recommender systems for ontology visualisation (Thellmann et al., 2015; Sváb-Zamazal et al., 2016). The advances in ontology visualisation have also prompted efforts to classify and categorise the extensive inventory of existing visualisation techniques. For instance, Katifori et al. (2007) define six categorisations of ontology visualisation techniques, including indented list, node-link and tree, zoomable, space-filling, focus with context or distortion, and 3D information landscapes, based on their presentation, interaction, functionality, and dimensions used. Dudáš et al. (2018) define five categorisations of ontology visualisation techniques, including 1.5D, 2D, 2.5D, 3D, and temporal dimensional visualisations that can be presented as indented lists and node-link diagrams with various layouts such as force-directed, tree, treemap, radial, circle, and Euler diagrams. A notable observation from these classifications is that there are two dominant techniques that are commonly employed in ontology visualisation, namely indented lists and node-link diagrams, which evidently serve as the foundation of many other renditions in ontology visualisation generation. In addition, a second observation is that a 'one-size-fits-all' approach has dominated most ontology visualisation tools, where the same visualisation is presented to

the user regardless of this person's information needs and visual preferences. This highlights a potential research opportunity to design adaptive visual systems that may adjust to the user needs in real time, i.e. as the person interacts with a given ontology visualisation. To this end, we employ the aforementioned two commonly used ontology visualisation techniques in a study that aims to predict user success ahead of task completion, in an effort to identify adaptation opportunities where potential visual adjustments may be necessary to improve user success.

Adaptive technologies have been well-studied in related fields such as information visualisation, where early systems (Casner, 1991; Mackinlay, 1986) have investigated adaptations based on a given task or data properties known a priori, with successful results indicating expedited user performance during visual information processing. Further studies in dynamic visualisation adaptations such as those discussed in Grawemeyer (2006) provide visualisation recommendations based on inferred user expertise and preferences from earlier interactions, with evaluation results showing more accurate and efficient user performances. Since prior interaction data may not always be available in every scenario, later research (Gotz and Wen, 2009) has investigated the feasibility of adaptation based on real-time user monitoring, such as data patterns generated from mouse clicks. However, direct user input such as mouse clicks may not always be available in every interactive scenario. For instance, if the user is simply gazing at a visualisation, it would be difficult to provide adaptation in that scenario. To overcome such issues, recent research has focused on adaptation models based on alternative data sources including non-invasive physiological sensors such as eye tracking. In cognitive and perceptual psychology, eye tracking has long been established as a suitable means to quantify user attention patterns in reading and search activities (Rayner, 1995) as well as information processing (Rayner, 1998). In human-computer interaction, eye tracking has also been applied to various research topics such as identifying pattern variances in different visualisations (Goldberg and Helfman, 2011), types of tasks (Iqbal and Bailey, 2004), and cognitive efforts required of the user (Plumlee and Ware, 2006). As eye trackers become increasingly affordable over the years, coupled with their ubiquitous integration in gaming (Velloso and Carter, 2016) and application in computer accessibility for people with physical disabilities (Betke et al., 2002), it can be envisioned that gaze-based adaptations may soon be part of mainstream visualisation systems. In the field of ontology visualisation, eye tracking has been used as an evaluation tool to compare the usability of common visualisation techniques (Fu et al., 2017). Preliminary research using gaze data, mouse clicks, and event logs to infer user success has shown promising results in predictive analytics (Fu and Steichen, 2019; Fu et al., 2019). However, these approaches are designed to support post hoc adaptations and are dependent on direct user input that may not always be available or possible as discussed earlier. Thus, in an effort to extend these prior research efforts, this paper aims to investigate the benefits of utilising gaze data generated from an interactive session that is still in progress. More specifically, we show that it is feasible to predict a user's success with a given task before the session is concluded. It is envisioned that these predictions would in turn inform the underlying visual system to adjust/add visual cues as appropriate, in an attempt to increase user success and improve the user experience for the remainder of an interactive session. We recognise that it may be useful to determine in some scenarios if one ontology visualisation may be more optimal than another for a given user, although arguably, such a prediction would only allow intrusive adaptations, such as changing the visualisation altogether while a user is already deeply engaged in a

task. As such, this research thrives for less intrusive adaptations to an ongoing interaction between the user and an ontology visualisation, whereby the existing visualisation may be accompanied by, e.g. adaptive overlays, and highlighted or modified axes, as proposed in Steichen and Fu (2019).

### 3 Experiment design

We designed and executed a controlled eye tracking experiment to simulate a human-ontology interaction scenario, whereby participants were asked to evaluate a set of class mappings between a given ontology pair assisted by their visualisations. In order to evaluate whether mappings are correct and complete, the participants must interact with visualisations of the given ontologies in order to understand the semantics within. As such, this experimental design provides the necessary means to investigate gaze-based predictive analytics.

We used two different datasets (two ontology pairs and their respective mapping gold standards) taken from two tracks at the ontology alignment evaluation initiative (OAEI), namely the Conference Track and the Biomedical Ontologies (BioMed) Track<sup>1</sup>. Both tracks cover general knowledge in their respective domains. More specifically, the Conference ontologies present broad knowledge of typical conference organisations, and the BioMed ontologies are a subset taken from the original dataset (containing over 70,000 classes) presenting knowledge of the human anatomy. Both ontology pairs used in the experiment are similar in size, with approximately 200 classes in total. For example, a participant was shown a pair of BioMed ontology visualisations and a list of mappings between this ontology pair at the same time, and was asked to answer true/false questions such as whether mapping *Organism* from the source ontology to *Metabacteria* in the target ontology is correct. In addition, the participant was asked to enter the ontology classes they believe should have been mapped together but not already included in the given mapping list. Participants evaluated class mappings exclusively in the experiments, and the visualisations illustrated solely *is-a* relationships. Inspired by precision, recall, and f-measure commonly used in mapping evaluation, we define a participant's correctness success as this person's accurate identifications of correct mappings in a given set. Likewise, we define a participant's completeness success as this person's correct additions of new mappings to the given set. Finally, we define a participant's overall success as this person's combined score that includes both correctness and completeness success. Correctness, completeness, and overall success scores range between 0 and 1, where the closer the value is to 1 the higher a person's success.

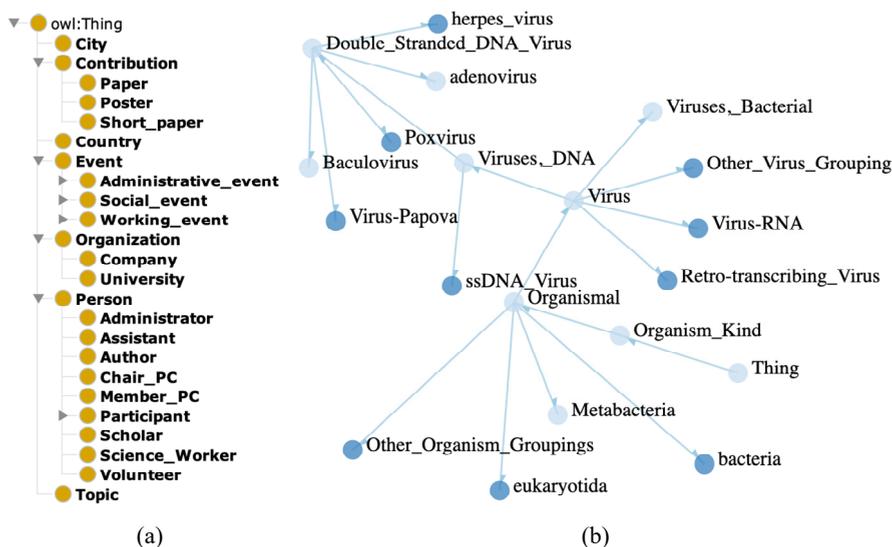
The visualisations and the ontologies used in the experiment are not intended to be exhaustive, but rather as example scenarios to provide the necessary environment for human-ontology interaction that is needed to validate the concept proposed in this paper. Though the aforementioned example scenarios are sufficient for the purpose of validating the feasibility of user success predictions based on gaze data, it is necessary to note that there are other potential scenarios and examples of user tasks involving ontology visualisations. As such, the findings reported in this paper shall be understood within the scope of the scenarios demonstrated in the experimental trials.

The mapping sets given to the participants were generated based on the OAEI gold standards, where we included incorrect and incomplete mapping results to allow scoring of user success as discussed earlier. In addition, we recorded the time on task, i.e. the

length of time it took a participant to complete a given mapping evaluation task. In both domains, the participants would need to identify the same number of correct and incorrect mappings, and to add the same number of missing mappings (each task scenario requires a participant to identify 13 correct results, 3 incorrect results, and to add 7 new mappings) in order to achieve 100% success score. This design aims to minimise potential bias that a participant's time on task is not driven by the number of mappings to be evaluated or to be created.

To demonstrate that user success predictions may be achieved across different visualisation techniques, we employed two types of commonly used ontology visualisations as identified previously in related work, namely indented list and node-link diagrams as shown in Figure 1. An example visualisation of a conference ontology using the indented list visualisation is shown in Figure 1(a), where super-subclass relationships are illustrated by indentations with arrow directions indicating whether classes can be toggled. Another example visualisation of a Biomedical ontology using node-link diagram is shown in Figure 1(b), where class relationships are illustrated by connected edges, with darker nodes indicating classes that can be further expanded to reveal additional child nodes. We used the D3 JavaScript Library<sup>2</sup> to generate the node-link diagrams and Protégé<sup>3</sup> to generate the indented list visualisations in our experiment, where the participants were instructed to only interact with the visualisations themselves but not any other features in Protégé.

**Figure 1** Ontology visualisation techniques investigated in the controlled experiment, (a) indented list visualisation of a conference ontology (b) node-link visualisation of a biomedical ontology (see online version for colours)



Finally, we used a Tobii 2150 eye tracker with a frame rate of 50Hz against a 21.3" monitor with 1600×1200 pixels resolution to capture gaze data. Each recording session began with calibration to ensure maximum tracking validity. No additional head mounts or sensors were used in our experiment setup, which thereby closely resembles common interactions of average users in the real world. Additionally, the eye tracker is able to tolerate small head movements and user blinking, although participants were encouraged

to minimise large movements once a recording had begun. We seated participants on stationary chairs without swivels or wheels to facilitate successful recordings of eye gaze data. A total of 36 participants took part in our experiment, who were undergraduate and graduate students majoring in Computer Science, Biomedical, Mechanical, and Electrical Engineering. Overall, we counterbalanced task orders, domains, and visualisations, whereby a participant interacted with the same domain or visualisation type only once, and different task orders were issued across all participants. This design therefore aims to minimise ordering and learning effects that may potentially impact on users' success or time on task.

**Table 1** Raw eye gaze data generated for a participant in one recording session

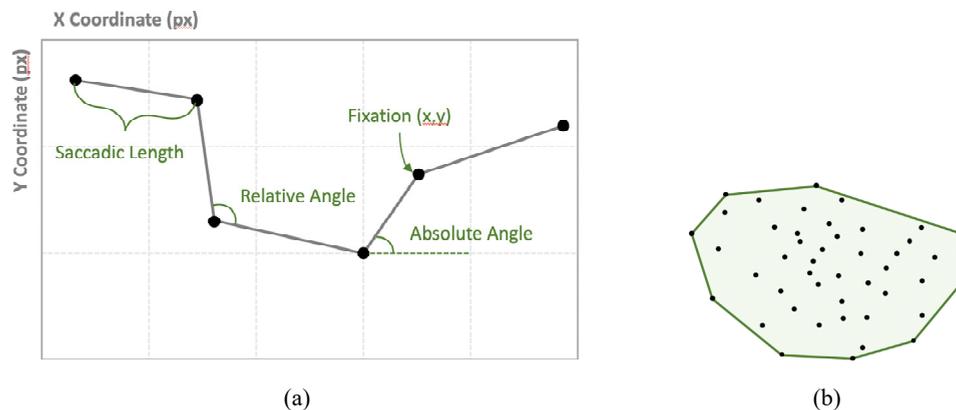
<i>Gaze measure</i>	<i>Description</i>
Fixation count	The number of fixation points found in an interaction
Fixation duration	The duration of a fixation, measured in milliseconds
Saccade count	The number of saccades found in an interaction
Saccade duration	The duration of a saccade, measured in milliseconds
Saccade length	The distance between fixations, measured in pixels
Saccade-to-fixation ratio	The sum of saccadic duration divided by the sum of fixation duration
Scanpath duration	The duration of all fixations and saccades, measured in milliseconds
Scanpath length	The distance of all saccades found in an interaction, measured in pixels
Convex hull	The size of an area defined by the bounding fixations found in an interaction, measured in square pixels
Absolute saccadic angle	The gradient of a saccade with respect to the horizontal axis, measured in degrees
Relative saccadic angle	The gradient of two consecutive saccades, measured in degrees
Pupil dilation	The widening of the pupils, measured in millimetres

Based on a set of raw gaze data produced by the Tobii eye tracker, we generated several gaze measures as shown in Table 1. Fixations indicate moments when a user's gaze is relatively stationary, e.g. when a person fixates on a point of interest on the screen. An example fixation is shown in Figure 2(a). During an eye tracking recording, there are many moments where a person's eyes are relatively still, and the total count of such fixations can be generated after a recording session. A low Fixation count may be understood as an indicator for effective displays of information, i.e. if given two visualisations representing the same amount of information, a lower fixation count may entail fewer visual foci needed for the user to perceive the same information (Goldberg and Kotval, 1999). Each fixation is also associated with a duration, where a shorter fixation duration in a recording may indicate overall faster information extractions and consumptions by the user (Goldberg and Kotval, 1999). Saccades refer to the quick eye movements between pairs of fixations and are associated with count and duration similar to fixations. In addition, we can measure the length of a saccade in pixels as the distance between two fixations – an example saccadic length is shown in Figure 2(a). Saccadic lengths may be interpreted as rapid redirections of users' attention, where longer saccadic lengths may indicate fewer unnecessary interim fixations (Goldberg and Kotval, 1999). Given a user's total fixation and saccade durations, we can then generate this person's

saccade-to-fixation ratio or search-to-process ratio in other words. These ratios provide an overall impression of how users have divided their time during an interaction, i.e. whether they spent more time on searching or on processing the given visual information.

The complete sequence of fixations and saccades is known as *scanpath*, which is associated with duration (the sum of all fixation and saccade durations) as well as length (the sum of all saccadic lengths in an eye tracking recording). Given two visualisations representing the same information and supporting the same tasks, it is likely that the more efficient visualisation with superior visual support would lead to generally shorter scanpath durations for the user (Goldberg and Kotval, 1999). However, it is also important to perceive scanpath durations in the context of its corresponding *convex hull*, which represents the smallest bounding area containing all fixations in the scanpath. An example of convex hull is shown in Figure 2(b). In addition, as a user repeats the process of searching and processing information during an interaction, there are various directional changes throughout a scanpath that may be measured via *relative and absolute angles*. Relative angles measure the saccadic degree between a pair of consecutive fixations, and absolute angles measure the degree of a saccade with respect to the horizontal axis, as demonstrated in Figure 2(a). Saccades with relative angles greater than  $90^\circ$  are typically understood as backtrack indicators, suggesting a possible change in the user's goal or mismatched visual representations and the user's expectations (Goldberg and Kotval, 1999). Lastly, *pupil dilations* may indicate the cognitive workload demanded by a user during an interaction (Poole and Ball, 2005), and in particular, increases in pupil sizes entail more cognitive effort (Marshall, 2000; Pomplun and Sunkara, 2003). During calibration, we collected baseline pupil sizes from each participant. We then collected their pupil sizes during interactions with the given ontology visualisations and mapping tasks. These pupil sizes collected during task completions were then compared to their respective baselines in order to generate pupil dilations for each participant.

**Figure 2** Descriptive gaze measures generated from raw eye tracking data, (a) an example of fixation, saccade, saccadic length, absolute and relative angles (b) an example of convex hull (see online version for colours)



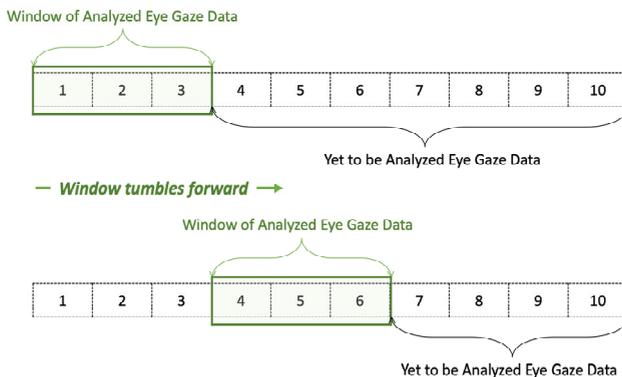
Finally, the gaze metrics shown in Table 1 can be categorised based on the type of activities they are perceived to measure. More specifically, fixation and saccade count, saccadic length, scanpath length, and convex hull, may be interpreted as measures of *information search* activities, whereby an effective visualisation is likely to direct users'

attention rapidly to relevant visual cues exclusively with optimal scanpath required in a given area of interest while minimising unnecessary intermediate steps. Likewise, average fixation duration and saccade-to-fixation ratio may be interpreted as measures of *information processing* activities, whereby an effective visual display is likely to support rapid information extraction in addition to faster and fewer searches of relevant visual cues. Furthermore, pupil dilation, absolute and relative saccade angles may be interpreted as measures of *cognitive workload*, whereby an effective visualisation is likely to demand less cognitive effort from users and provide helpful visual cues consistently aligned with user goals and expectations leading to minimised backtracks.

#### 4 Classification experiments using tumbling window predictive analysis

We conducted classification experiments using the Waikato Environment for Knowledge Analysis (WEKA) machine learning toolkit (version 3.8.2) (Frank et al., 2016) to predict, using the above gaze measures, each user’s likely success when evaluating a given set of mappings. As previously discussed, we generated three types of success for each user, namely correctness success, completeness success, and overall success. Using a median split for each type of success scores, the classification task was to infer whether a user is likely to belong to the high (i.e. above the median success score) or low (i.e. below the median success score) group for that success measure. We used 10-fold cross-validation (stratified) for model evaluation, and experimented with a total of 44 different classifiers with default parameter configurations, and the most notable results of selected classifiers are reported in this paper. The predictions are compared to those results generated using a baseline classifier, namely *ZeroR*, which predicts the majority based on the target. *ZeroR* is commonly used as a benchmark that provides a necessary point of reference when determining a baseline performance, thus it is sufficient for the purpose of this study.

**Figure 3** Data analysis using the tumbling window technique (see online version for colours)



Since we are interested in investigating the feasibility of predicting a user’s success before the person completes a given task, the classification experiments do not use the gaze data for the complete user session. Instead, the experiments use a *tumbling window predictive analysis*, which aims to focus on a series of fixed-sized and non-overlapping continuous frames in the eye tracking data (see Figure 3 for an illustration of a tumbling window). Specifically, given a stream of contiguous gaze data, the analysis window starts

with the first few lines of data (such as those numbered as 1, 2, and 3), makes predictions based on this data, and then tumbles forward to make subsequent predictions based on the subsequent data. The goal of such a tumbling window technique is to focus on the most current gaze behaviour of an individual user when predicting this person's success. In particular, as a user interacts with ontology visualisation, this person may experience various visual needs such as searching and processing information, assessing whether the given information is relevant, redirecting attention and visual focus as needed etc., which are likely to be reflected in this person's gaze data generated throughout various stages during an interaction. By focusing on current snapshots of users' gaze data, the tumbling window technique aims to produce the most up-to-date predictions with continuous updates of the last known gaze data for each user, which may differ from those generated at earlier stages for that person.

The window size of the tumbling window may be implemented in a number of ways. In this paper, we demonstrate two example methods, using either percentiles or absolute time intervals. First, the *tumbling window by percentage* method defines the window size as a fixed percentage of the full gaze dataset. For instance, the analysis window may begin with the first 20% of a person's gaze data to make predictions, and tumble to the next 20th percentile of the gaze dataset to make another prediction, where this process is repeated in the subsequent predictions until the 100th percentile is reached to make the final prediction. There are many variations of percentiles that can be applied to implement a percentile-based tumbling window, in this paper, we use a 10% interval to demonstrate this proposed method in predictive analytics. Such an implementation of the tumbling window may be appropriate in scenarios where adaptations are made post interaction in retrospect, with the goal to influence future user interactions with a given visual system, since percentiles are not yet known in advance during interactions in real time. For example, it may be desirable for a visual system to identify those notable percentiles during an interactive process that persistently display usage issues, which may be best examined in the presence of complete gaze datasets post hoc.

An alternative implementation of the window size may be defined with a fixed time frame using the *tumbling window by absolute time* method. For instance, the analysis window may generate predictions based on the gaze data collected from the first 5 minutes of interaction, it then tumbles forward to the next 5 minutes, and so on, until the user completion of the task. In this paper, we use a 2-minute interval as an example to demonstrate the proposed implementation of a tumbling window by absolute time in gaze analytics. Such an implementation of the tumbling window may be appropriate in scenarios where ad hoc adaptations are desirable, such as a necessity in adapting to users during their interactions with a visual system. Finally, the implementations of the tumbling window and their respective results presented in this paper are not intended to be exhaustive, since there are many other possible instrumentations of the tumbling window, but rather, we aim to demonstrate that predictive user success analyses based on gaze data is not only feasible, but also may provide a first step towards adaptive ontology visualisation.

## 5 Results

This section presents classification results generated using different input gaze features sets, i.e. user success predictions that are inferred based on the complete set of gaze

features (Section 5.1), or subsets indicating user activities such as information search (Section 5.2), information processing (Section 5.3), or cognitive workload (Section 5.4) as previously discussed in Section 3. We demonstrate that depending on the feature set used, some classifiers have shown additional improvements in various classification tasks, such as higher accuracies or even sooner predictions. As discussed in Section 4, while we experimented with many more classifiers, only results of those notable classifiers are reported in this paper. It is therefore necessary to note that the figures presented in the remainder of this section contain selected classifiers in a given configuration that may vary across trials. These selected classifiers are not an exhaustive list of all classification models that have outperformed the baseline, but are intended to demonstrate example usage of a tumbling window in predictive user analyses based on gaze data.

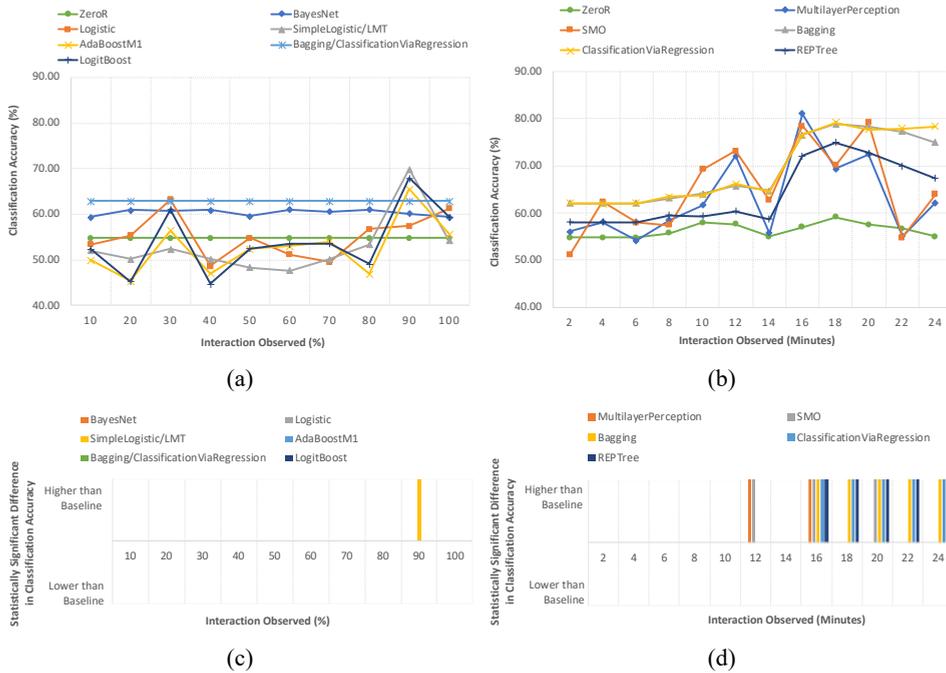
### 5.1 *Trial 1 – predictions using complete gaze feature set*

The classification results generated using the *tumbling window by percentage* method when predicting a user’s likely success in evaluating whether a given mapping set is correct or not are shown in Figure 4(a) and Figure 4(c). The baseline classifier achieved an accuracy of 54.76% throughout various percentiles, which were outperformed by several other classifiers and a few notable examples are shown in Figure 4(a). The highest accuracy was achieved by the *SimpleLogistic* and the logistic model tree (*LMT*) classifiers at the 90th percentile with an accuracy of 69.93%, which was statistically significant compared to the baseline classifier at the same percentile [Figure 4(c)]. There appears to be two percentiles in particular, namely the 30th and 90th percentiles, where relative peak accuracies were generated by a number of classifiers. These results suggest that early predictions of users may be possible, although more accurate predictions may be achieved at later stages of the interaction when inferring users’ likelihood to assess correct mappings in the given scenario.

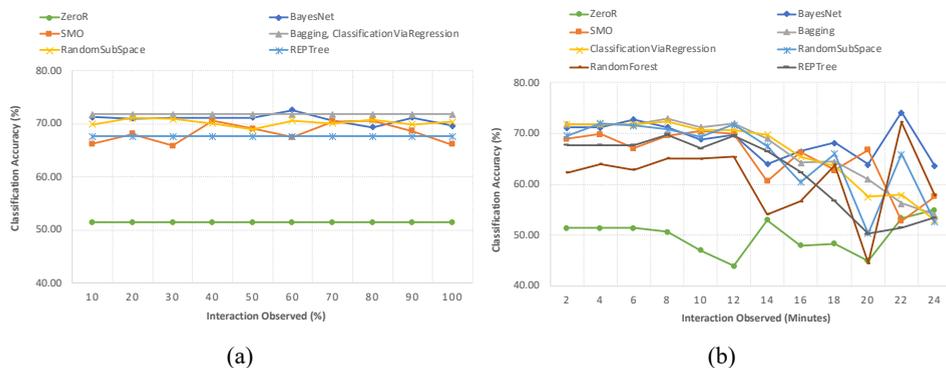
Using the *tumbling window by absolute time* method, we generated classification results for users’ correctness success as shown in Figure 4(b) and Figure 4(d). The baseline classifier generated its peak accuracy of 59.17% after 18 minutes into the interaction [Figure 4(b)]. This is outperformed by a number of classifiers as early as after 2 minutes into the interaction. The highest accuracy of 81.15% was generated by the *MultilayerPerceptron* classifier after 16 minutes. Furthermore, though a number of higher accuracies were reported during early time intervals, the number of statistically significant results peaked at 16 minutes [Figure 4(d)]. These findings suggest that in the given scenario, there may be an ideal time interval to produce the most accurate predictions on users’ correctness success with the highest statistical power, though it may be possible to detect the same trend during early stages of the interaction in some cases.

When predicting users’ success using the *tumbling window by percentage* method when evaluating whether a set of mappings are complete or not, the baseline classifier generated an accuracy of 51.43% throughout all percentiles. A number of classifiers [Figure 5(a)] generated higher accuracies and with statistically significant results at all percentiles [Figure 5(c)]. The highest accuracy was produced by the *BayesNet* classifier with 72.67% accuracy at the 60th percentile. This finding further indicates that it is not only possible to predict users’ completeness success throughout all stages of the interaction, but it is also feasible to generate accurate predictions as early as the first 10th percentile in the given scenario.

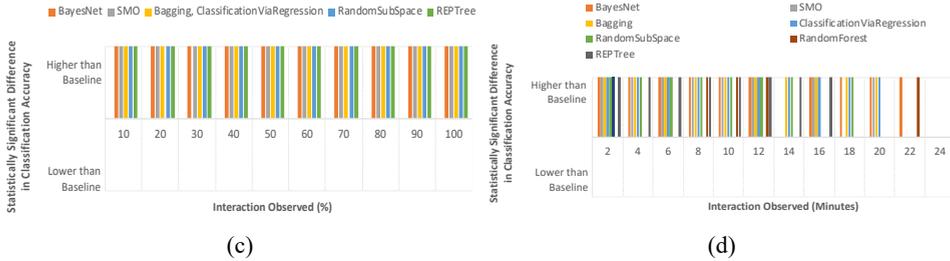
**Figure 4** Classification results of users' correctness success based on complete gaze datasets using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



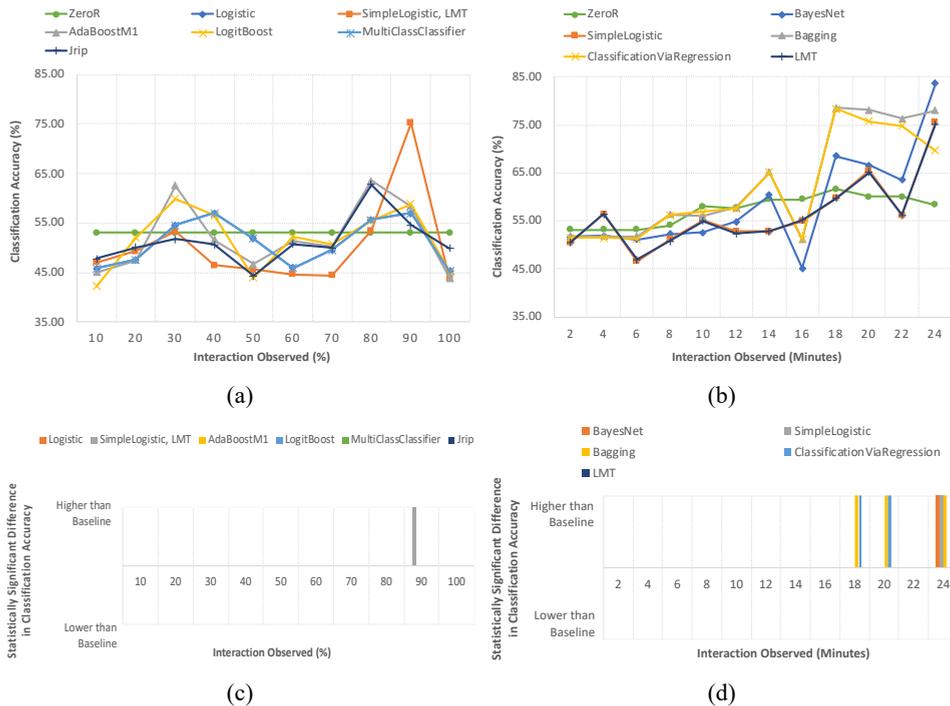
**Figure 5** Classification results of users' completeness success based on complete gaze datasets using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



**Figure 5** Classification results of users' completeness success based on complete gaze datasets using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (continued) (see online version for colours)



**Figure 6** Classification results of users' overall success based on complete gaze datasets using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



The classification results generated for completeness success predictions using the *tumbling window by absolute time* method are presented in Figure 5(b) and Figure 5(d). The highest accuracy generated by the baseline classifier was 55% after 24 minutes of interaction [Figure 5(b)]. This was outperformed by a number of classifiers as early as

after 2 minutes with statistically significant results [Figure 5(d)]. Most notably, a peak accuracy at 74.17% was generated by the *BayesNet* classifier after 22 minutes of interaction. Furthermore, a higher number of statistically significant results were reported during earlier time intervals, e.g. from the beginning until approximately 12 minutes of interaction, compared to later time intervals, e.g. after approximately 20 minutes of interaction. These findings suggest that there are not only multiple opportunities to make accurate predictions regarding a user's completeness success in the given scenario, but such predictions can also be achieved after just a few minutes into an interaction.

When predicting users' overall success in the given task using the *tumbling window by percentage* method, the baseline classifier generated an accuracy of 53.10% at all percentiles [Figure 6(a)], which were also achieved by a number of other classifiers such as the *NaiveBayesMultinomialText* classifier, the *SGDText* classifier, the *CVParameterSelection* classifier, the *MultiScheme* classifier, the *Stacking* classifier, the *Vote* classifier, the *WeightedInstancesHandlerWrapper* classifier, and the *InputMapped* classifier. The highest accuracy of 75.24% was produced by the *SimpleLogistic* and the *LMT* classifiers at the 90th percentile, which was statistically significantly more accurate than that of the baseline [Figure 6(c)]. Relative peak accuracies appeared approximately around the 30th, 80–90th percentiles, which is consistent with the aforementioned findings of correctness success predictions. These results suggest early inferences of users' overall success may be achieved in some cases, offering additional motivation to further validate other implementations of the tumbling window technique.

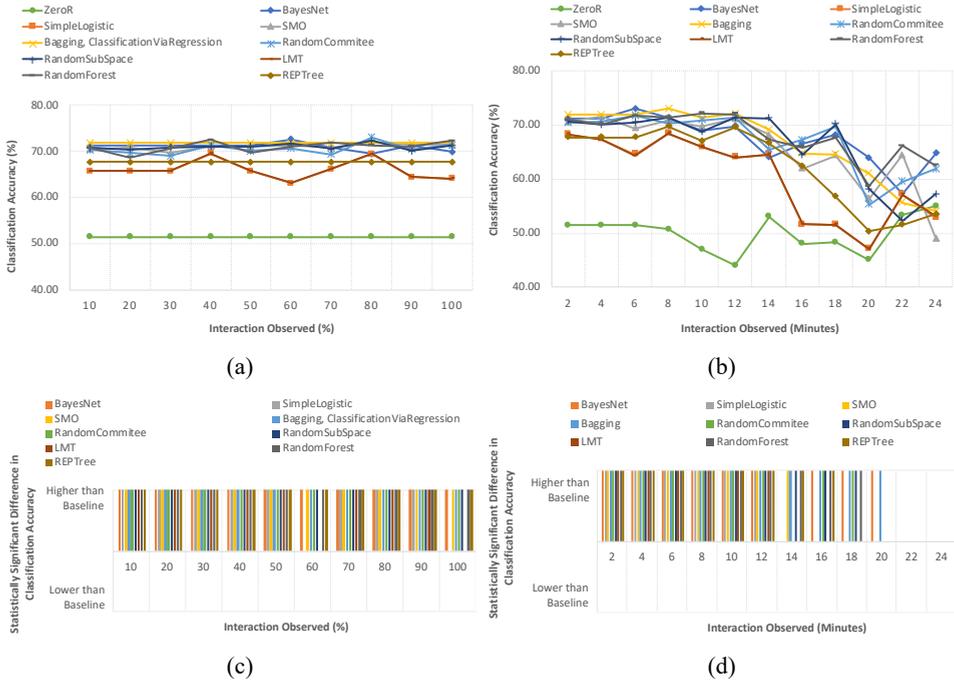
Figure 6(b) shows the classification results generated for users' overall success prediction using the *tumbling window by absolute time* method. The baseline classifier achieved its peak accuracy of 61.67% at the 18-minute interval. Several classifiers generated higher accuracies, where the highest accuracy of 83.67% was generated by the *BayesNet* classifier after 24 minutes of interaction. Relatively higher accuracies were reported by a number of classifiers after approximately 4 minutes of interaction and again after 18 minutes of interaction. Different from the predictions on users' correctness and completeness success, several classifiers' predictions on the users' overall success were less accurate at 16 minutes compared to those of the baseline classifier. The earliest statistically significant results that were more accurate than those of the baseline were determined after 18 minutes by the *bagging* and the *classification via regression* classifier [Figure 6(d)]. These findings suggest that early accurate predictions on users' overall success may be possible, though more accurate results with higher statistical power are likely to be produced towards later stages of the interaction in the given scenario.

## 5.2 Trial 2 – predictions using gaze feature subset indicating information search activities

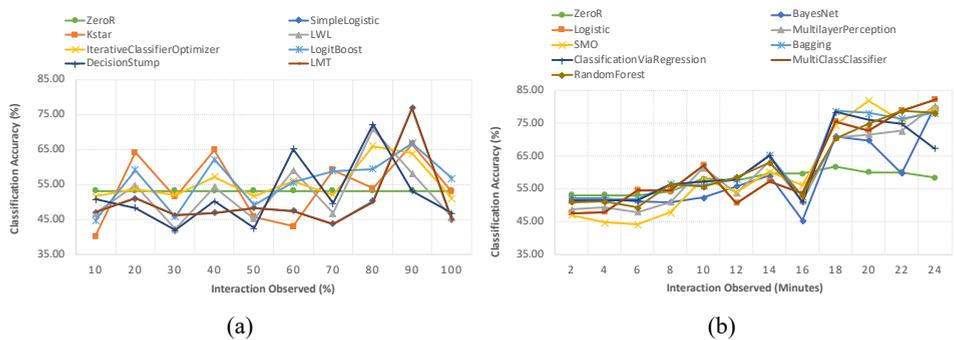
The predictions on users' correctness success based on gaze measures that indicate information search activities are shown in Figure 7. The baseline classifier generated an accuracy of 54.76% using both the *tumbling window by percentage* method and the *tumbling window by absolute time* method. Higher accuracies are shown in Figure 7(a) and Figure 7(b), where a peak accuracy of 71.74% was achieved by the *LogitBoost* classifier using the *tumbling window by percentage* method at the 90th percentile and a peak accuracy at 79.17% was generated by the *ClassificationViaRegression* classifier using the *tumbling window by absolute time* method after 18 minutes into the interaction.



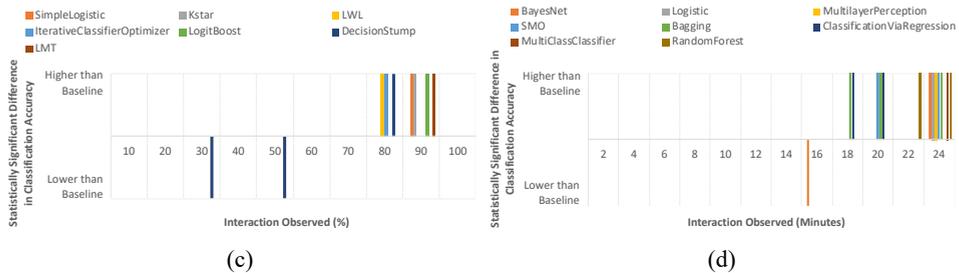
**Figure 8** Classification results of users' completeness success based on gaze feature subsets indicating information search using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



**Figure 9** Classification results of users' overall success based on gaze feature subsets indicating information search using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



**Figure 9** Classification results of users’ overall success based on gaze feature subsets indicating information search using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (continued) (see online version for colours)



When predicting users’ overall success, the baseline classifier generated an accuracy of 53.10% using the *tumbling window by percentage* method and a peak accuracy at 61.67% after 18 minutes using the *tumbling window by absolute time* method. These results were outperformed by a number of classifiers with statistically significant accuracies as shown in Figure 9. A peak accuracy at 76.79% was achieved by the *SimpleLogistic* and the *LMT* classifier using the *tumbling window by percentage* method at the 90th percentile [Figure 9(a)], and a peak accuracy at 82.17% was generated by the *Logistic* classifier after 24 minutes using the *tumbling window by absolute time* method [Figure 9(b)]. In both tumbling window methods, statistically significant predictions were found towards the later stages of the analysis as shown in Figure 9(c) (after the 80th percentile) and Figure 9(d) (after 18 minutes of interaction). Overall, top-performing classifiers generally achieved higher accuracies as well as a greater number of statistically significant predictions using the *tumbling window by absolute time* method.

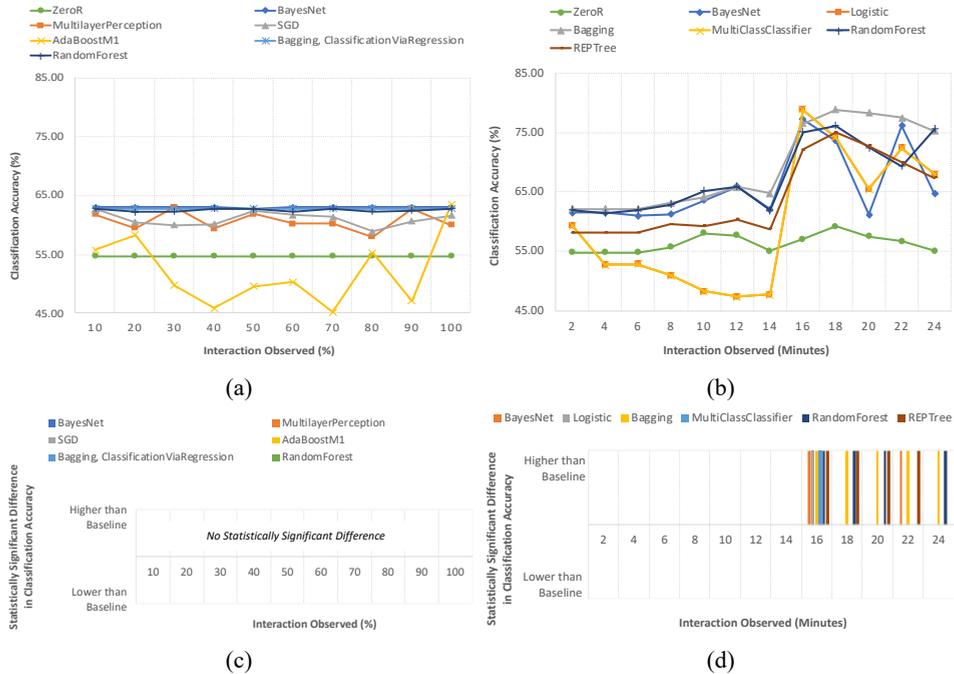
### 5.3 Trial 3 – predictions using gaze feature subset indicating information processing activities

When predicting users’ correctness success using gaze feature sets indicating information processing activities, the baseline classifier achieved an accuracy at 54.76% using the *tumbling window by percentage* method [Figure 10(a)], and a peak accuracy at 59.17% after 18 minutes of interaction using the *tumbling window by absolute time* method [Figure 10(b)]. These predictions were outperformed by a number of classifiers in both implementations of the tumbling window. Most notably, higher accuracies as well as a greater number of statistically significant predictions were found when using absolute time to define window size [Figure 10(d)], where a peak accuracy at 78.85% was generated after 18 minutes of interaction using the *Logistic* classifier.

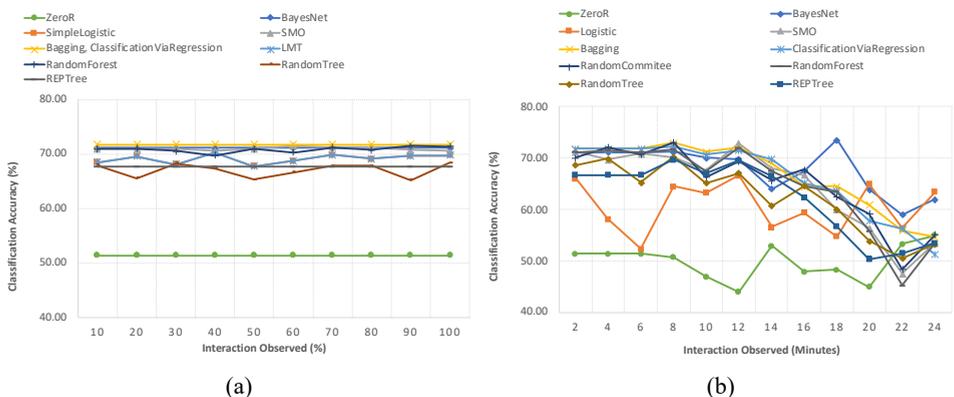
When predicting users’ completeness success based on gaze features indicating information processing activities, the baseline classifier achieved an accuracy of 51.43% using the *tumbling window by percentage* method [Figure 11(a)] and a peak accuracy of 55% after 24 minutes of interaction using the *tumbling window by absolute time* method [Figure 11(b)]. A number of classifiers generated higher accuracies with statistically significant results as shown in Figure 11(c) and Figure 11(d), where the highest accuracy found was 77.40% after 12 minutes of interaction using the *tumbling window by absolute*

time method, where a greater number of statistically significant predictions were also found towards the beginning of the interaction.

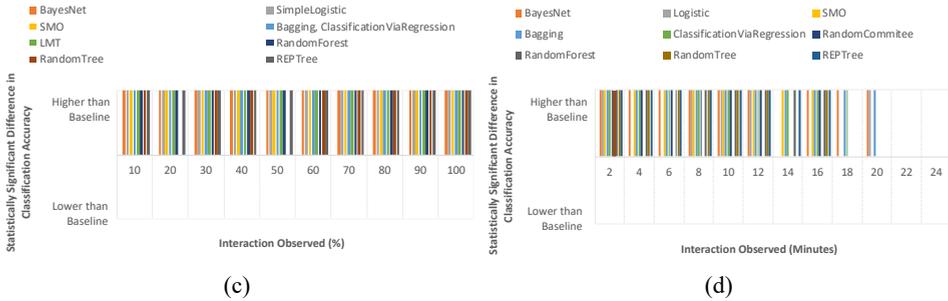
**Figure 10** Classification results of users' correctness success based on gaze feature subsets indicating information processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



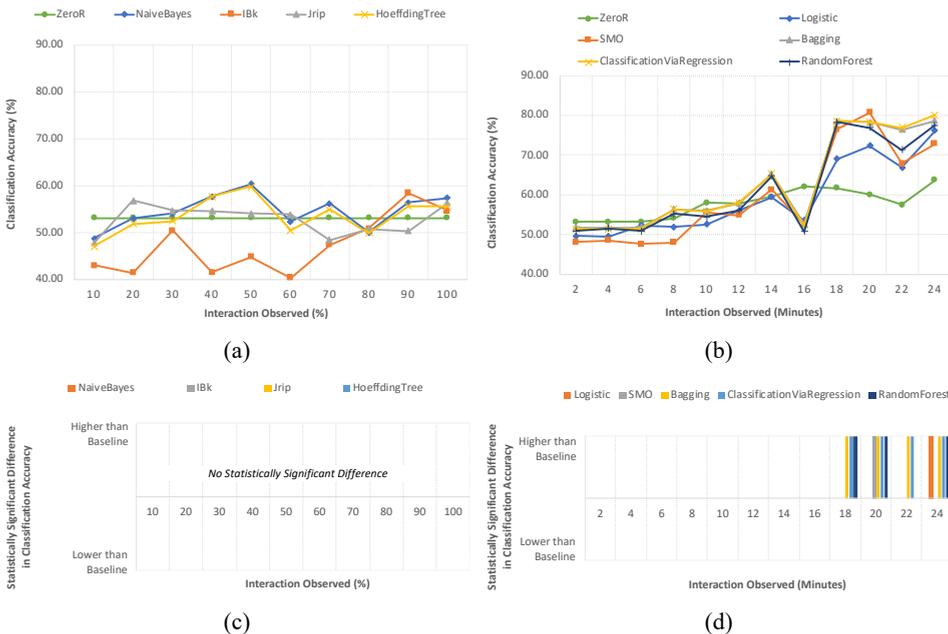
**Figure 11** Classification results of users' completeness success based on gaze feature subsets indicating information processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



**Figure 11** Classification results of users' completeness success based on gaze feature subsets indicating information processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (continued) (see online version for colours)



**Figure 12** Classification results of users' overall success based on gaze feature subsets indicating information processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



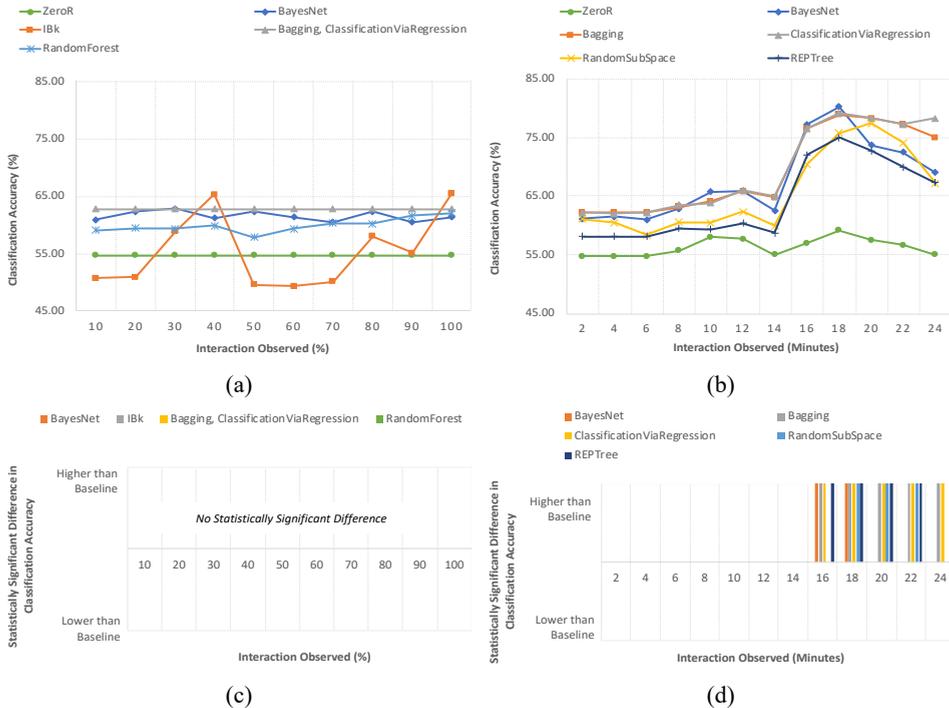
When predicting users' overall success using gaze features indicating information search activities, the baseline classifier generated an accuracy of 53.10 % using the *tumbling window by percentage* method [Figure 12(a)] and a peak accuracy of 63.67% after 24 minutes of interaction using the *tumbling window by absolute time* method [Figure 12(b)]. These results were outperformed by several classifiers using the *tumbling*

window by percentage method, although these predictions were not statistically significant. Better results were generated using the *tumbling window by absolute time* method, where the highest accuracy achieved was 80.67% after 20 minutes of interaction using the sequential minimal optimisation (*SMO*) classifier. In addition, we found an increased number of statistically significant predictions towards the later stages of interactions using the *tumbling window by absolute time* method [Figure 12(d)].

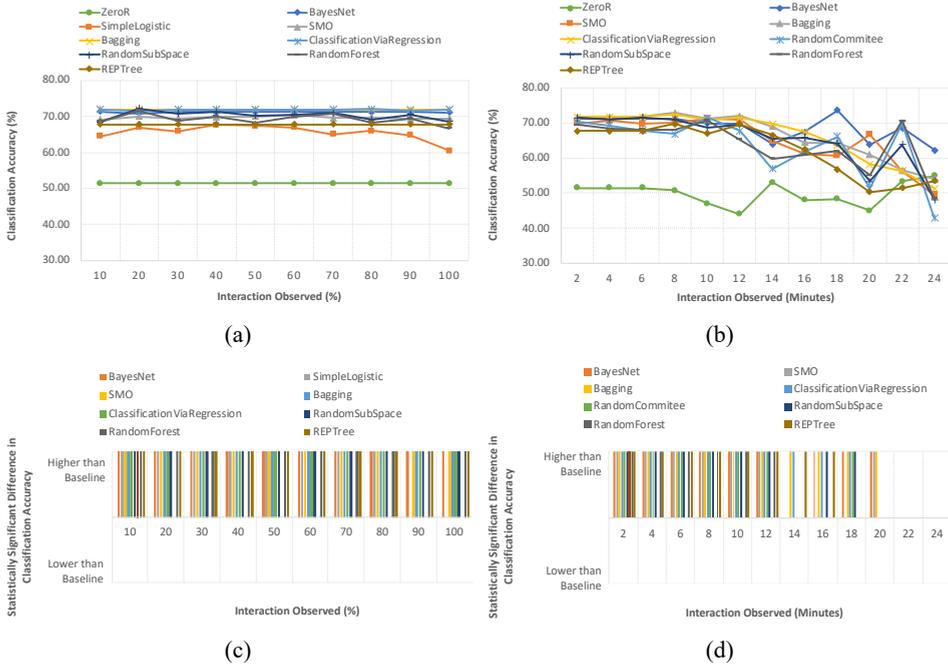
### 5.4 Trial 4 – predictions using gaze feature subset indicating cognitive workload

The predictions on users’ correctness success using gaze features indicating cognitive workload are shown in Figure 13. The baseline classifier generated an accuracy of 54.76% using the *tumbling window by percentage* method [Figure 13(a)] and a peak accuracy of 59.17% after 18 minutes of interaction using the *tumbling window by absolute time* method [Figure 13(b)]. These results were outperformed by several classifiers in both implementations of the tumbling window method, with statically significant improvements achieved using absolute time to define the tumbling windows [Figure 13(d)]. A notable number of statistically significant results were generated toward later interaction stages, where the highest accuracy generated was 80.33% after 18 minutes.

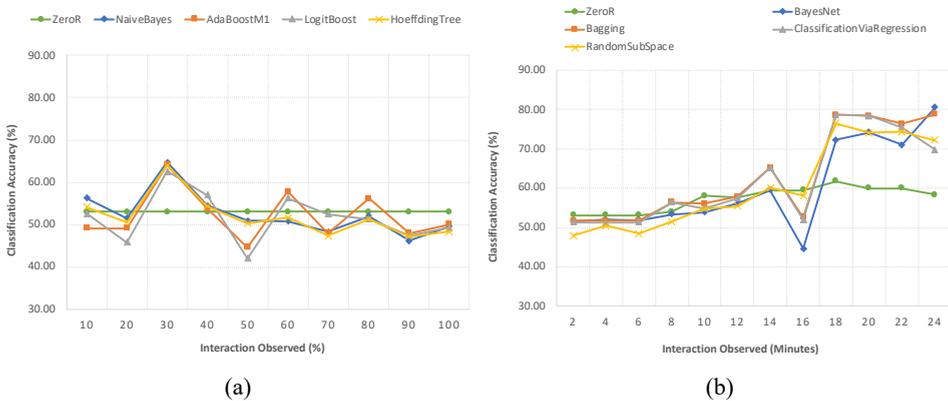
**Figure 13** Classification results of users’ correctness success based on gaze feature subsets indicating cognitive workload processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



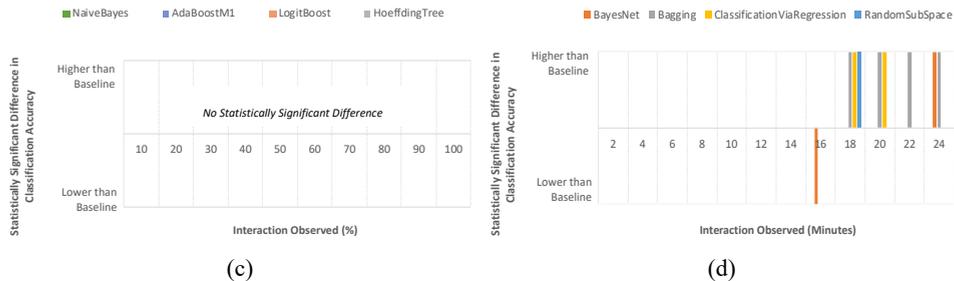
**Figure 14** Classification results of users' completeness success based on gaze feature subsets indicating cognitive workload processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



**Figure 15** Classification results of users' overall success based on gaze feature subsets indicating cognitive workload processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (see online version for colours)



**Figure 15** Classification results of users' overall success based on gaze feature subsets indicating cognitive workload processing using two implementation methods of the tumbling window, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (c) statistical significance of predictions using tumbling window by percentage (d) statistical significance of predictions using tumbling window by absolute time (continued) (see online version for colours)



When predicting users' completeness success, the baseline classifier achieved an accuracy of 51.43% using the *tumbling window by percentage* method [Figure 14(a)] and a peak accuracy of 55% after 24 minutes using the *tumbling window by absolute time* method [Figure 14(b)]. A number of classifiers achieved better results that are statistically significant as shown in Figure 14(c) and Figure 14(d). Also, a notably increased number of statistically significant results were reported early on during an interaction regardless of the specific implementation to define the size of the tumbling window. The highest accuracy achieved was 73.58% after the 18 minutes of interaction using the *BayesNet* classifier and the *tumbling window by absolute time* method.

When predicting users' overall success based on gaze features indicating cognitive workload, the baseline classifier generated an accuracy of 53.10% using the *tumbling window by percentage* method [Figure 15(a)] and a peak accuracy of 61.67% after 18 minutes using the *tumbling window by time* method [Figure 15(b)]. These results were outperformed by a number of classifiers with statistically significant differences most notably at later stages of the interaction, where absolute time was used to define tumbling window size [Figure 15(d)]. The highest accuracy achieved was 80.50% after 24 minutes.

### 5.5 Peak predictions & influential gaze features

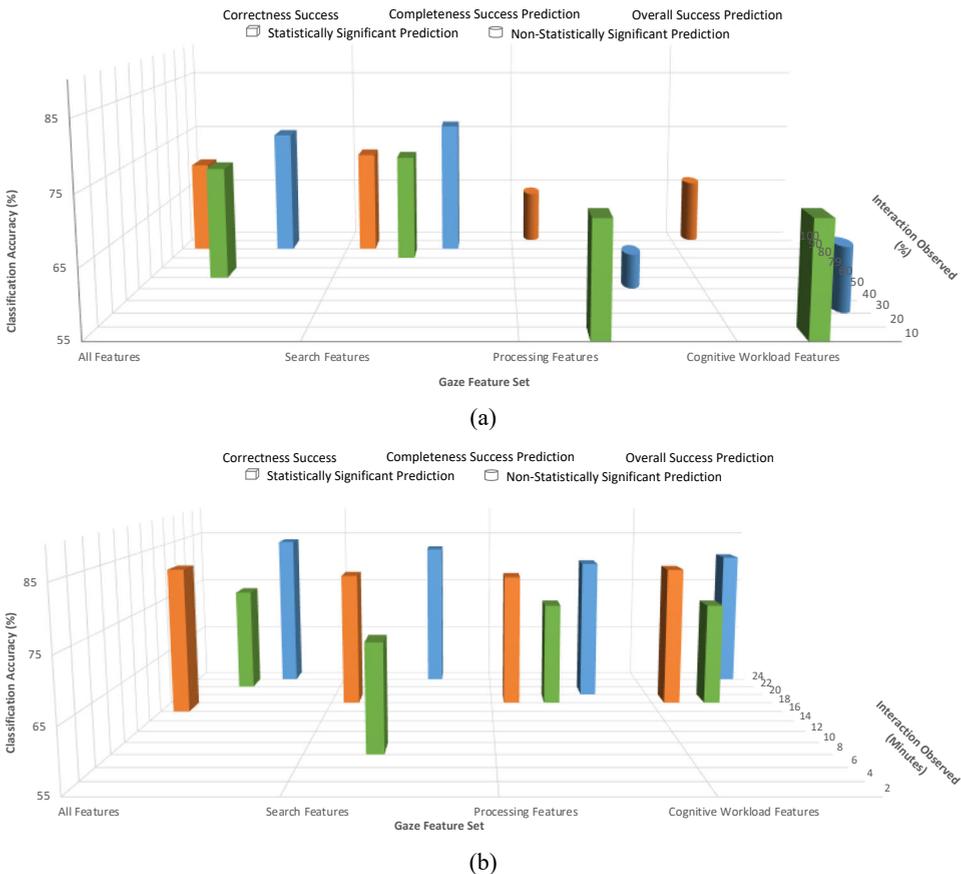
A summary of peak accuracies achieved across all trials when predicting the three types of user success are shown in Figure 16, where the *x-axis* indicates the feature sets used in the predictive analysis, the *y-axis* indicates the accuracy of the predictions, and the *z-axis* indicates the percentiles or the time intervals used to define the tumbling window size.

In the *tumbling window by percentage* method [Figure 16(a)], not all peak accuracies found were shown to be statistically significant. The earliest statistically significant accuracies were generated at the 10th percentile when predicting users' completeness success using the feature subsets indicating either processing or cognitive workload activities. The majority of peak accuracies were found at later percentiles across all other types of success predictions and feature set combinations. When predicting users' correctness success, the highest accuracy found was 71.74% using the *LogitBoost* classifier and the feature subset indicating search activities at the 90th percentile. When predicting users' completeness success, the highest accuracy found was 72.67% using the

*BayesNet* classifier and the complete gaze feature sets at the 60th percentile. When predicting users' overall success, the highest accuracy found was 76.79% using the *NaiveBayes* classifier and the feature subset indicating processing activities at the 90th percentile.

In the *tumbling window by absolute time* method (Figure 16b), all peak accuracies were shown to be statistically significant. The earliest peak accuracy was found after 8 minutes when predicting users' completeness success using the feature subset indicating users' search activities. In addition, peak accuracies achieved by various classifiers were found at a number of time intervals throughout the interaction. Furthermore, we found that accuracies also increased as more time passed in the interaction. When predicting users' correctness success, the highest accuracy found was 81.15% using the *MultilayerPerceptron* classifier and the complete feature set after 16 minutes. When predicting users' completeness success, the highest accuracy found was 74.17% using the *BayesNet* classifier coupled with either the complete feature set or the feature subset indicating search activities after 12 minutes. When predicting users' overall success, the highest accuracy found was 83.67% using the *BayesNet* classifier and the complete feature set after 24 minutes.

**Figure 16** Peak classification results found across gaze feature sets, (a) predictions using tumbling window by percentage (b) predictions using tumbling window by absolute time (see online version for colours)



In addition to examining overall classifier accuracies, we also investigated in more detail the most influential gaze features based on the classification models that have achieved the highest accuracies, as well as the directionality of these features, i.e. whether a higher or lower value for this feature leads to high or low user success. As shown in Table 2, a combination of several gaze features were associated with high user success, such as less dispersed and longer average fixation durations, less dispersed and larger relative saccadic angles, more dispersed but shorter average saccadic lengths, and greater changes in pupil dilation. These results suggest that users who are more successful at the given tasks are likely to have directed their attention with greater changes, searched for relevant visual cues positioned both nearby and further apart, and consistently spent more time to fixate on a visual cue once it is found. Furthermore, in the case of completeness success using the tumbling window by time method, a distinct directional change was found for the pupil dilation feature. More specifically, at the beginning of the interaction (before approximately 8 minutes), higher changes in pupil dilation were found to be associated with higher user success. However, this directionality would shift to the opposite after 8 minutes, where higher pupil dilations were found to be associated with lower user success throughout the remainder of an interactive session. This finding suggests that if the user was not able to comprehend the visual cues at the beginning of an interaction (indicated by enlarged pupil dilation), it is unlikely that this person will overcome visual challenges despite continued efforts in later interaction. This result further emphasises on the importance of timely interventions and highlights early adaptive visual cues as potential solutions in the process of better support for the user.

**Table 2** Most influential features to predict user success

<i>Feature</i>	<i>Directionality</i>
mean <i>Fixation Duration</i>	high value → <i>high success</i>
standard deviation of <i>Fixation Duration</i>	low value → <i>high success</i>
mean <i>Relative Saccadic Angles</i>	high value → <i>high success</i>
standard deviation of <i>Relative Saccadic Angles</i>	low value → <i>high success</i>
mean <i>Saccadic Length</i>	low value → <i>high success</i>
standard deviation of <i>Saccadic Length</i>	high value → <i>high success</i>
<i>Pupil Dilation Change</i>	high value → <i>high success</i>

## 6 Conclusions and future work

This paper aims to demonstrate the feasibility of gaze-based predictions in the context of human-ontology interaction, where the predictive analytics presented in this paper are envisioned to inform the underlying ontology visualisation system to potentially intervene either ad hoc or post hoc in an effort to improve user success and enhance user experience. This research aims to advance the design and development of adaptive ontology visualisation systems in their strategies to tailor to an individual user. If it is possible to correctly predict whether a user is likely to succeed or fail (step 1), we can then investigate the appropriate means to intervene (step 2). The scope and motivation of this paper falls within step I, where the findings aim to inform the realisation of step 2. The experimental trials demonstrate several implementations of a tumbling window

designed to achieve ad hoc and post hoc user predictions. The experiment results show that user success may be inferred with significantly higher accuracies compared to those of a baseline classifier at various stages of the interaction with two commonly used ontology visualisation techniques. In addition, we demonstrate that accurate user predictions can be achieved early and throughout interactive stages, suggesting that there may be ample opportunities to potentially intervene with alternative visual suggestions or visualisation alterations to better support the user. Furthermore, different gaze feature subsets are compared using several configurations, coupled with variations of a tumbling window to identify most influential gaze features. This section summarises the conclusions and lessons learned.

- 1 *Observation 1:* accurate gaze-based user predictions are possible in the most commonly used ontology visualisation techniques including indented lists and node-link visualisations. This finding is encouraging as it demonstrates that predictive analytics remain accurate while being visualisation-agnostic, since correct predictions were generated regardless of the specific ontology visualisation used by the user in our trials. This result highlights the approach presented in this paper is not visualisation specific and may be transferrable to other ontology visualisation techniques. Future research could investigate additional ontology visualisation techniques such as those included in Katifori et al. (2007) and Dudáš et al. (2018), as well as comparing alternative predictive strategies to determine the most appropriate analytics for a given ontology visualisation.
- 2 *Observation 2:* accurate user predictions can be generated despite user and domain variations. The experimental trails have demonstrated successful predictions in the presence of varied user backgrounds and task domains. Considering a mixture of prior knowledge in a given domain is likely to be inevitable in a real-world setting, the results shown in this paper are highly motivating in the proposed gaze-based predictive approach, where user predictions remain accurate despite possible variations of user and domain variables. Future research could include additional variables in the experiments to potentially identify whether distinct user groups (e.g. novice vs. expert users) may influence prediction accuracies. Likewise, there are many other scenarios that can be further explored in future research to compare predictive differences based on the given ontological task (e.g. entity search vs. entity creation). Furthermore, future research may explore potential influences of notable user characteristics [e.g. cognitive styles (Steichen et al., 2020; Steichen and Fu, 2020), perceptual and spatial abilities (Conati et al., 2014), domain and visualisation expertise (Toker et al., 2012)] on prediction accuracies.
- 3 *Observation 3:* the accuracies of the predictions as well as the number of statistically significant accuracies generated using the *tumbling window by absolute time* method are consistently higher than those produced using the *tumbling window by percentage* method, suggesting the former method may be more appropriate when defining the tumbling window size in the given task scenario. Although, it may be argued that the percentiles used in the trials may not have been optimal, and it is not yet known if further improved accuracies using a different percentile may be possible. Future research could investigate alternative strategies to determine the most optimal percentiles to apply in a given scenario. Likewise, although better results were generated using a 2-minute interval, it is unknown whether shorter or

longer time intervals may have produced even more accurate results. Future experiments could focus on developing automated algorithms to best identify the most appropriate time intervals that would lead to the highest possible accuracies in a given task scenario. Moreover, future predictions may be classified by additional dimensions such as the type of visualisation or the particular order of the tasks, to determine whether there may be further improvements of accuracies.

- 4 *Observation 4*: earlier predictions of users' completeness success were found in a number of experimental trials compared to other types of user success. This finding suggests that inferring whether users will be successful at evaluating the completeness of a mapping set may be achieved much sooner than predicting their likely success at evaluating the correctness of the same mapping set. This result indicates that the differences may be grounded in the nature of these two types of evaluation tasks, and that the visual needs may be intrinsically different prompting distinctive gaze activities and the consequent predictions thereof. However, these speculations require further experiments to validate. Future research could focus on investigating whether certain visual needs demand diverse approaches to generate user predictions and adaptations in the process of achieving personalised ontology visualisations. Moreover, it may be beneficial to investigate other definitions of window-based predictive analysis techniques. For example, research in Steichen et al. (2014) demonstrates accurate user predictions using eye gaze data collected over time. Future research could investigate whether accumulated windows may lead to statistically significant accuracies when predicting the user success.
- 5 *Observation 5*: it may be beneficial to utilise partial gaze data to achieve earlier predictions or higher accuracies in some cases. For example, in the *tumbling window by percentage* method, the earliest statistically significant predictions of user's completeness success were reported at the 10th percentile (an accuracy of 71.83% using the *Bagging* and the *ClassificationViaRegression* classifier) using gaze data subsets indicating either users' processing activities or cognitive workload. Also, the highest accuracy generated when predicting users' overall success (an accuracy of 76.79% using the *SimpleLogistic* and the *LMT* classifier) was achieved by using gaze subsets that indicated users' search activities. These findings suggest that depending on the goal of the predictive analysis, gaze subsets may be sufficient to produce more desirable results compared to the complete gaze dataset. In other words, an adaptive system may be able to achieve its highest accuracies with less computational overhead if desired.
- 6 *Observation 6*: an investigation of the most influential gaze features indicates users who were successful at the given tasks tend to demonstrate fewer indications of undirected visual searches. There is evidence to suggest that greater user success is often associated with assertive changes in their visual searches and focused attention once a relevant visual cue is found. In particular, increased cognitive workload at the earlier stage of an interaction may be perceived as a positive indicator for user success, compared to later stages where continued increase was found to be associated with user failure. This finding further supports the rationale to adapt to individual users early in view of potentially preventing user failures.

Lastly, while we have demonstrated the feasibility of employing gaze data in user success predictions coupled with a tumbling window analysis technique, we have not investigated how these predictions can be integrated in an adaptive ontology visualisation system. Future research focusing on designing and evaluating subsequent processes in realising adaptations in the form of user intervention or alternative visualisation recommendation would be necessary to advance the body of knowledge in adaptive ontology visualisation. In particular, future studies may exclusively assess the extent to which users' success can be improved utilising inferred predictions towards achieving adaptive ontology visualisation. Moreover, other forms of physiological data may be used in future experiments to complement gaze data to provide potentially more accurate or earlier detections of likely user success or failure. For example, it may be beneficial to utilise electroencephalogram (EEG) responses from users that may provide supplementary data when predicting user success.

## References

- Bartolomeo, S.D., Pepe, G., Savo, D.F. and Santarelli, V. (2018) 'Sparqling: painlessly drawing SPARQL queries over graphol ontologies', *VOILA@ISWC*.
- Betke, M., Gips, J. and Fleming, P. (2002) 'The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities', in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 10, No. 1, pp.1–10, March, doi: 10.1109/TNSRE.2002.1021581.
- Burch, M. and Lohmann, S. (2015) 'Visualizing the evolution of ontologies: a dynamic graph perspective', *VOILA@ISWC*.
- Casner, S.M. (1991) 'Task-analytic approach to the automated design of graphic presentations', *ACM Transactions on Graphics*, April, Vol. 10, No. 2, pp.111–151.
- Chen, C-M. (2008) 'Intelligent web-based learning system with personalized learning path guidance', *Journal of Computers & Education*, Vol. 51, No. 2, pp.787–814, Elsevier.
- Cho, Y.H., Kim, J.K. and Kim, S.H. (2002) 'A personalized recommender system based on web usage mining and decision tree induction', *Journal of Expert Systems with Applications*, Vol. 23, No. 3, pp.329–342, Elsevier.
- Conati, C., Carenini, G., Hoque, E., Steichen, B. and Toker, D. (2014) 'Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making', *Computer Graphics Forum*, June, Vol. 33, No. 3, pp.371–380.
- Dudáš, M., Lohmann, S., Svátek, V. and Pavlov, D. (2018) 'Ontology visualization methods and tools: a survey of the state of the art', *The Knowledge Engineering Review*, Vol. 33, E10.
- Frank, E., Hall, M.A. and Witten, I.H. (2016) 'The WEKA workbench', *Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Morgan Kaufmann.
- Fu, B. and Steichen, B. (2019) 'Using behavior data to predict user success in ontology class mapping – an application of machine learning in interaction analysis', in *Proceedings of the IEEE 13th International Conference on Semantic Computing (ICSC)*, pp.216–223.
- Fu, B., Noy, N.F. and Storey, M-A. (2017) 'Eye tracking the user experience - an evaluation of ontology visualization techniques', *Semantic Web Journal – Interoperability, Usability, Applicability*, Vol. 8, No. 1, pp.23–41, IOS Press.
- Fu, B., Steichen, B. and McBride, A. (2020) 'Tumbling to succeed: a predictive analysis of user success in interactive ontology visualization', in *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics (WIMS 2020)*, pp.78–87.
- Fu, B., Steichen, B. and Zhang, W. (2019) 'Towards adaptive ontology visualization – predicting user success from behavioral data', *International Journal of Semantic Computing*, Vol. 13, No. 4, pp.431–452.

- Goldberg, J. and Helfman, J. (2011) 'Eye tracking for visualization evaluation: reading values on linear versus radial graphs', *Inf. Vis.*, July, Vol. 10, No. 3, pp.182–195.
- Goldberg, J.H. and Kotval, X.P. (1999) 'Computer interface evaluation using eye movements: methods and constructs', *International Journal of Industrial Ergonomics*, Vol. 24, No. 6, pp.631–645,
- Gotz, D. and Wen, Z. (2009) 'Behavior-driven visualization recommendation', in *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI'09)*, pp.315–324.
- Grawemeyer, B. (2006) 'Evaluation of ERST: an external representation selection tutor', in *Proceedings of the 4th International Conference on Diagrammatic Representation and Inference (Diagrams'06)*, pp.154–167.
- Graziosi, A., Iorio, A.D., Poggi, F. and Peroni, S. (2017) 'Customised visualisations of linked open data', *VOILA@ISWC*.
- Haag, F., Lohmann, S., Siek, S. and Ertl, T. (2015) 'QueryVOWL: visual composition of SPARQL queries', *ESWC, LNCS*, Vol. 9341, pp.62–66.
- Iqbal, S.T. and Bailey, B.P. (2004) 'Using eye gaze patterns to identify user tasks', presented at the *Grace Hopper Celebration of Women in Computing*.
- Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C. and Giannopoulou, E. (2007) 'Ontology visualization methods - a survey', *ACM Computing Surveys*, ACM, New York, Vol. 39, No. 4, Article 10.
- Lambrix, P., Dragisic, Z., Ivanova, V. and Anslow, C. (2016) 'Visualization for ontology evolution', *VOILA@ISWC*.
- Lohmann, S., Negru, S., Haag, F. and Ertl, T. (2015) 'Visualizing ontologies with VOWL', *Semantic Web*, Vol. 7, No. 4, pp.399–419.
- Mackinlay, J. (1986) 'Automating the design of graphical presentations of relational information', *ACM Transactions on Graphics*, April, Vol. 5, No. 2, pp.110–141.
- Marshall, S. (2000) *Method and Apparatus for Eye Tracking and Monitoring Pupil Dilation to Evaluate Cognitive Activity*, July, U.S. Patent 6090051, USA.
- Motta, E., Mulholland, P., Peroni, S., d'Aquin, M., Gomez-Perez, J.M., Mendez, V. and Zablith, F. (2011) 'A novel approach to visualizing and navigating ontologies', *ISWC, LNCS*, Vol. 7031, pp.470–486.
- Ochs, C., Geller, J., Musen, M.A. and Perl, Y. (2017) 'Real time summarization and visualization of ontology change in Protégé', *VOILA@ISWC*.
- Plumlee, M.D. and Ware, C. (2006) 'Zooming versus multiple window interfaces: cognitive costs of visual comparisons', *ACM Trans Comput-Hum Interact*, June, Vol. 13, No. 2, pp.179–209.
- Pomplun, M. and Sunkara, S. (2003) 'Pupil dilation as an indicator of cognitive workload in human-computer interaction', in *Proceedings of HCI International*, Vol. 3, pp.542–546.
- Poole, A. and Ball, L.J. (2005) 'Eye tracking in human-computer interaction and usability research: current status and future', in Ghaoui, C. (Ed.): *Encyclopedia of Human-Computer Interaction*, Idea Group, Inc., Pennsylvania.
- Rayner, K. (1995) 'Eye movements and cognitive processes in reading, visual search, and scene perception', *Studies in Visual Information Processing*, pp.3–22.
- Rayner, K. (1998) 'Eye movements in reading and information processing: 20 years of research', *Psychol. Bull.*, November, Vol. 124, No. 3, pp.372–422.
- Sabol, V., Tschinkel, G., Veas, E., Hoefler, P., Mutlu, B. and Granitzer, M. (2014) 'Discovery and visual analysis of linked data for humans', *ISWC, LNCS*, Vol. 8796, pp.309–324.
- Santana-Pérez, I. (2018) 'Graphless: using statistical analysis and heuristics for visualizing large datasets', *VOILA@ISWC*.
- Steichen, B. and Fu, B. (2019) 'Towards adaptive information visualization - a study of information visualization aids and the role of user cognitive style', *Frontiers in Artificial Intelligence*, November, Vol. 2, pp.1–10, Article 22.

- Steichen, B. and Fu, B. (2020) ‘Cognitive style and information visualization-modeling users through eye gaze data’, *Frontiers in Computer Science*, Vol. 2, p.562290, doi: 10.3389/fcomp.2020.562290.
- Steichen, B., Ashman, H. and Wade, V. (2012) ‘A comparative survey of personalised information retrieval and adaptive hypermedia techniques’, *Information Processing and Management*, Vol. 48, No. 4, pp.698–724.
- Steichen, B., Carenini, G. and Conati, C. (2014) ‘Inferring visualization task properties, user performance, and user cognitive abilities from eye gaze data’, *ACM Transactions on Interactive Intelligent Systems*, July, Vol. 4, No. 2, Article 11.
- Steichen, B., Fu, B. and Nguyen, T. (2020) ‘Inferring cognitive style from eye gaze behavior during information visualization usage’, in *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pp.348–352.
- Sváb-Zamazal, O., Dudás, M. and Svátek, V. (2016) ‘Augmenting the ontology visualization tool recommender: input pre-filling and integration with the OOSP ontological benchmark builder’, *SEMANTiCS*.
- Theilmann, K., Galkin, M., Orlandi, F. and Auer, S. (2015) ‘LinkDaViz – automatic binding of linked data to visualizations’, *ISWC, LNCS*, Vol. 9366, pp.147–162.
- Toker, D., Conati, C., Carenini, G. and Haraty, M. (2012) ‘Towards adaptive information visualization: on the influence of user characteristics’, in Masthoff, J., Mobasher, B., Desmarais, M.C. and Nkambou, R. (Eds.): *User Modeling, Adaptation, and Personalization. UMAP, LNCS*, Vol. 7379, Springer, Berlin, Heidelberg, [https://doi.org/10.1007/978-3-642-31454-4\\_23](https://doi.org/10.1007/978-3-642-31454-4_23).
- Vargas, H., Buil-Aranda, C., Hogan, A. and López, C. (2019) ‘RDF explorer: a visual SPARQL query builder’, *ISWC, LNCS*, Vol. 11778, pp.647–663.
- Velloso, E. and Carter, M. (2016) ‘The emergence of EyePlay: a survey of eye interaction in games’, in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY ‘16)*, ACM, New York, NY, USA, pp.171–185, DOI: <http://dx.doi.org/10.1145/2967934.2968084>.
- Zainab, S.S., Saleem, M., Mehmood, Q., Zehra, D., Decker, S. and Hasnain, A. (2015) ‘FedViz: a visual interface for SPARQL queries formulation and execution’, *VOILA@ISWC*.

## Notes

- 1 See details of these tracks at <http://oaei.ontologymatching.org/2019/>.
- 2 <https://d3js.org>.
- 3 <http://protege.stanford.edu>.