Competing Models: Inferring Exploration Patterns and Information Relevance via Bayesian Model Selection

Shayan Monadjemi, Roman Garnett, and Alvitta Ottley

Abstract— Analyzing interaction data provides an opportunity to learn about users, uncover their underlying goals, and create intelligent visualization systems. The first step for intelligent response in visualizations is to enable computers to infer user goals and strategies through observing their interactions with a system. Researchers have proposed multiple techniques to model users, however, their frameworks often depend on the visualization design, interaction space, and dataset. Due to these dependencies, many techniques do not provide a general algorithmic solution to user exploration modeling. In this paper, we construct a series of models based on the dataset and pose user exploration modeling as a Bayesian model selection problem where we maintain a belief over numerous competing models that could explain user interactions. Each of these competing models represent an exploration strategy the user could adopt during a session. The goal of our technique is to make high-level and in-depth inferences about the user by observing their low-level interactions. Although our proposed idea is applicable to various probabilistic model spaces, we demonstrate a specific instance of encoding exploration patterns as competing models to infer information relevance. We validate our technique's ability to *infer exploration bias*, *predict future interactions*, and *summarize an analytic session* using user study datasets. Our results indicate that depending on the application, our method outperforms established baselines for bias detection and future interaction prediction. Finally, we discuss future research directions based on our proposed modeling paradigm and suggest how practitioners can use this method to build intelligent visualization systems that understand users' goals and adapt to improve the exploration process.

Index Terms—User Interaction Modeling, Bayesian Machine Learning

◆

1 INTRODUCTION

Visual analytics has transformed the process of reasoning with data by viewing humans and machines as teammates with unique strengths. Due to the growing complexity and volume of data, there is a need for more intelligent visual analytic systems to assist in data exploration, decision-making, and communication of findings. Studies suggest that even with interactive visualizations, there are human-related factors that hinder data exploration and informed decision-making. For instance, analysts often search through a large amount of information where the irrelevant portions of the data can be distracting and exceed the limits of human cognition [19]. An intelligent visual analytic system can respond by only showing portions of the data relevant to the analyst. An everyday user might interact with only a biased subset of the data which impedes decision-making [10]. An intelligent visual analytic system can respond by informing the user about their biases through notifications or mitigate their biases through modifying the visualization. Creating an effective intelligent visual analytic system requires careful design considerations including the level of intrusiveness and the means of intervention [42]. Regardless of what form this *intelligent visual analytic system* takes on, it must be able to offer well-informed assistance to the user.

One promising approach to informing useful machine response is through capturing and analyzing users' interaction data. Research in the area of *Analytic Provenance* has shown that interaction data can reveal valuable information about the user and their analysis strategies and how the machine teammates can better assist the user. For example, Dabek and Caban proposed a grammar-based approach to uncover common patterns among a group of users [7], resulting in suggestions to assist users in data exploration. Battle et al. [1] used a Markov chain that utilized navigation behavior to determine portions of a satellite image that a given user will likely explore in the future, resulting in improving latency by 430%. Similarly, Ottley et al. [33] and Wall et al. [41] used hidden Markov models to infer users' attention and

cognitive bias respectively, resulting in anticipating future clicks and analyzing user exploration bias. For a more comprehensive survey on user modeling with provenance data, see Ragan et al. [35] and Xu et al. [43]. Unfortunately, much of the existing solutions depend on the visualization encoding or the interaction design, and do not easily provide a general algorithmic solution to uncovering exploration patterns and quantifying information relevance.

In this paper, we take a more comprehensive view of user modeling that encompasses what the user has done, is doing, and will do in the future. Specifically, we use observed interactions as model evidence, and we leverage Bayesian model selection to reason about the subset of a multi-dimensional dataset and the collection of dimensions that most likely gave rise to the observations. This approach provides a valuable framework to build, maintain, and select models that we derive from the dimensions in the dataset, and it offers two key advantages. First, it utilizes straightforward probability theory that is easy to understand and implement. Second, a unique benefit of this approach is that we can make high-level inferences about a user's exploration pattern by passively observing low-level interactions with the data points. Once we identify the data dimensions that are most relevant, we demonstrate that we can use our models to achieve a variety of user modeling objectives: *infer exploration bias*, *predict future interactions*, and *summarize an analytic session*. We validate our proposed framework using three crowdsourced test datasets and we compare our algorithm's performance to existing baseline models. We show that our modeling technique outperforms established baselines in most cases. Furthermore, we discuss the possibilities this modeling paradigm opens for future investigations. We summarize our contributions as follows:

- *We introduce a straightforward approach to inferring multi-dimensional data exploration strategy in real-time.* We encode different exploration strategies as a set of models and utilizes Bayesian model selection to reason about which strategy is likely to give rise to past interactions with a point-based visualization.

- *We demonstrate our method's flexibility to address multiple aspects of user explorations modeling when applied to information relevance.* Specifically, we show that we can apply this framework to infer exploration bias, to predict future interaction, and to summarize an analysis session.

- *We validate the proposed framework using three crowdsourced user interaction datasets.* Our results indicate that, depending on the application, our method outperforms established baselines.

- *Shayan Monadjemi is with Washington University. monadjemi@wustl.edu.*
- *Roman Garnett is with Washington University. garnett@wustl.edu.*
- *Alvitta Ottley is with Washington University. alvitta@wustl.edu.*

2 BACKGROUND

Learning and modeling user behavior is a common goal in the Visual Analytics community [33, 35, 36, 41]. *Analytic Provenance*, which broadly refers to the research goal of tracking and modeling the analysis process, provides valuable information for inferring certain characteristics of exploration. Ragan et al. [35] categorize provenance into five categories: provenance of data, visualization, interactions, insights, and rationale. Furthermore, they categorize the primary purpose of tracking provenance data as *recall* [12, 28, 39], *replication* [8, 24, 29, 32], *action recovery* [9, 11, 18, 26, 38], *collaborative communication* [5, 13, 14, 30], *presentation* [20, 21], and *meta-analysis* [22, 34]. This paper aims to leverage *interaction provenance* data to learn about the users and their goals in order to improve the visualization system.

2.1 User Modeling in Visualization

Recent work has begun to investigate how learning about the users' analytic process can help us improve visualization systems. For example, Battle et al. [1] demonstrated this concept by using user exploration modeling to intelligently pre-fetch data for a map visualization. They observed user actions such as {*hover, click, pan-left, pan-right, ...*}, maintained a Markov chain model, and ranked the data for pre-fetching according to the likelihood of the users taking actions corresponding to said data. Their approach improved the overall latency by 430%.

Brown et al. [3] have proposed *Dis-function* as a method to learn expert knowledge through interaction data as a distance function representing the underlying similarities between data points. In their framework, the expert user interacts with the visualization by positioning similar points close to one another. By observing the user's organization of points, the algorithm infers a distance function that mimics the experts' knowledge. In similar work, Iwata et al. [23] took a step further to minimize the number of user interactions needed for finding the desired visualization representing high dimensional data. Using an active learning approach, they queried the user to relocate points strategically in order to reduce the uncertainty of the visualization.

Dabek and Caban [7] have proposed a grammar-based approach to model the sequence of low-level interaction in a session. Their work addresses the overwhelming number of controls and interaction channels available on modern visual analytics systems by using finite state machines to make suggestions for future interactions during a session. While their technique is task-dependent and relies on training data from past users performing the same tasks, their findings show that appropriate recommendations help users in data exploration.

This set of investigations provide examples on how user exploration modeling can improve user experience and decision making outcomes in visual analytic systems. However, many of them suffer from dependence on visualization design, tasks, or datasets. In line with our motivation of building intelligent visual analytic system who observe user interactions and infer what the user has done, is doing, and will do in the future, we propose an algorithmic and generalizable approach to building competing models on how users may interact with data. Moreover, we use Bayesian iterative updating to maintain our beliefs during sessions. We specifically design our models to take a comprehensive view on interactive sessions by inferring exploration bias, predicting future interactions, and summarizing sessions. In the remainder of this section, we discuss the most closely related work to this paper.

2.2 Inferring Exploration Bias

In the context of visual analytics, *bias* may arise from different sources. Related literature has studied the phenomena of bias from the lens of *data, models, and users*. Gotz et al. [16] proposed a toolkit to combat selection bias, where the source of bias is in the *data*. For instance, filtering a real estate data set to include houses less than $300K will automatically introduce a bias towards the location of selected houses since price and neighborhoods are highly correlated. Cabrera et al. [4] proposed a visual analytic tool, FairVis, to combat bias in machine learning *models*. For example, a model designed for recommending employment candidates for interviews may unfairly disregard candidates solely due to their historic under-representation in a career field.

In this work, we are interested in bias from the lens of *users*. As the most relevant work to this paper, Wall et al. [41] introduced metrics to measure cognitive bias through user interaction with visualized data. Cognitive bias is measured in terms of how much of the dataset has been explored, which attributes of the data points were important to the user, and what value of those attributes have influenced the users' decisions. As an example, imagine a parent who is interacting with a map of crimes to determine if a specific neighborhood is safe for their family. They may begin by looking for sex crimes in that specific neighborhood. The algorithm soon determines that the *location* and *type* of the crime are the two attributes driving the user's exploration, while the *time* attribute has been uniformly explored. Presenting this bias to the user during the session may result in them confirming their interest in a subset of data, or it may encourage more exploration and mitigate unintentional bias. While mitigating information bias requires careful design considerations which are beyond the scope of this paper (e.g. degree of guidance, level of intrusiveness, and means of interference [42]), we demonstrate our technique's ability to infer intersectional bias in exploration which provides critical information for bias mitigation.

2.3 Predicting Future Interactions

The term *interaction* can take on numerous meanings in the context of interactive visualizations. For example, *interaction* may refer to low level events such as clicks, hovers, and drag/drops, or it may refer to higher level tasks such as filtering and sorting. As a result, Gotz et al. [16] have proposed a taxonomy where analytic behavior is categorized in to four groups: *tasks, sub-tasks, actions, and events*. Their taxonomy created a spectrum based on semantics, where *tasks* on one end of the spectrum refer to interactions with high semantics (e.g. identify market insights for promising investments) and *events* on the other end of the spectrum refer to interactions with poor semantics (e.g. an individual click).

In the current literature, *predicting future interactions* can take on different meanings as well. For example, Dabek and Caban [7] build models to predict future *low-level events*, whereas Ottley et al. [33] and Battle et al. [1] build models to predict which *data point* the user is likely to interact with next. Our work is related to the latter category, as we will infer the relavance of data points to the user in light of past visited data points and compare our technique with Ottley et al. [33] as a baseline. Ottley et al. [33] proposed a hidden Markov model (HMM) approach to maintain a belief over users' evolving attention and actions in a visualization system. They encode user clicks as a sequence of visual attributes with an added bias metric to determine what attribute is influencing the exploration. As user clicks arrive, they update the model and use particle filtering to infer a set of top-*k* point candidates for the next click. In contrast to their work, our models are built on the *data* rather than the *visual elements*. Since not all attributes of a multi-dimensional dataset can have a visual representation, we believe that building the model on the dataset itself opens up an opportunity for learning about users' interest in *any* of the attributes, not only the ones visualized.

2.4 Summarizing Analytic Sessions

The body of work in this area ranges from primitive undo/redo functionalities to communicating provenance data from collaborative sessions. Heer et al. [20] explore tools for visualizing provenance data, and discuss how such tools can be used to improve user interface design. Xu et al. [44] propose *Charts Constellations*, a visualization tool to aggregate provenance data from multiple analysts in order to communicate the depth of exploration in different parts of a dataset. Chen et al. [6] observe low-level interactions and create annotations to assist users with insight externalization. Gratzl et al. [17] propose *CLUE*, a modeling framework that joins exploration and communication of discoveries by automatically storing provenance data during exploration and generating a session report.

In another related work, Sarvghad et al. [37] designed an experiment where users explore a multi-dimensional dataset to gain insight into business performance of an online retailer. Their study suggests that

Table 1. Table of notations used in this paper and their descriptions.

Notation	Description
$\mathscr{D}$	The set of all data points visualized
$\mathscr{C}$	The set of data points interacted with
$\mathscr{X}$	The space of all possible data points
$\mathbb{M}$	The set of all models
$\mathscr{M}_i$	An individual model from $\mathbb{M}$
$\mathbb{A}_{c,i}$	The set of continuous attributes relevant to $\mathscr{M}_i$
$\mathbb{A}_{d,i}$	The set of discrete attributes relevant to $\mathscr{M}_i$

those who were provided with a summary on which attributes they have covered in their exploration were able to formulate more questions and explore more broadly without sacrificing the depth of insights. Our proposed model in this paper allows us to accomplish a similar premise as prior work: to summarize a session. This capability can simply communicate our model's understanding of the user or it can be presented to the user to promote more exploration in a similar fashion as Sarvghad et al. [37].

3 PRELIMINARY DEFINITIONS

We begin with some definitions to frame the reader's understanding of the problem space. We assume the user interacts with a set of point-based visual data and each point of interest $\vec{x}$ can be described by a vector of d continuous or discrete attributes. We define a corresponding d-dimensional **data space** of possible realizations of these attributes, $\mathscr{X} = \{X_1 \times X_2 \times ... \times X_d\}$, where X_i is the domain of the ith attribute. We will consider a visualization of a **dataset** – that is, a collection of points in the data space – $\mathscr{D} = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_m\}$, $\vec{x}_i \in \mathscr{X}$. We assume that a user generates a stream of **interaction data**, a sequence of items in the dataset the user has interacted with, which we will notate with $\mathscr{C} = \{\vec{c}_1, \vec{c}_2, ...\}$, $\vec{c}_j \in \mathscr{D}$. While our definition of interaction data is independent of the type of interaction, this paper considers examples where the interactions occur by *clicking* or *hovering*. Our technique, however, may be expanded to include datapoints from other means of interaction subject to limitations discussed in Section 8. We approach user interaction modeling via the framework of Bayesian model selection. We construct competing high-level models of a user's exploration pattern based on different attributes of $\mathscr{D}$, and use a given set of interaction data to determine which of these models is the most plausible. Here, a model is a parametric family of probability distributions to explain a given set of observations, and we will construct models that allow for a variety of different subspaces/subsets of the data space to be relevant (or not) for a user, spanning a large space of hypotheses. In the Bayesian approach, there are three major steps to probabilistic inference:

- The **prior** represents a belief distribution over possible values of the parameter before data is observed.
- The **likelihood** specifies the probability of observing a given set of observations (such as a click stream) assuming given values for the parameters.
- The **posterior** is the updated distribution over possible values of the parameter in light of the prior beliefs and the information in the observations.

Going forward, we initiate the *prior* belief and maintain the *posterior* belief over a set of competing models in Section 4.1. Moreover, we discuss the *likelihood*, or the relevance of data points in light of past observations, in section 4.2. Finally, we use the competing models to detect exploration bias, predict future interactions, and summarize a session (Section 5).

4 COMPETING MODELS FRAMEWORK

The novelty of this work is that we present different exploration patterns as models, and use the Bayesian model selection framework to maintain a belief over all possible exploration patterns. Our approach to user interaction modeling assumes that users interact with visualizations with some (possibly subconscious) pattern in mind. The ultimate goal of our technique is to detect and model these exploration patterns. We define

model space to be a finite set of models representing various exploration patterns. The model space is denoted by $\mathbb{M} = \{\mathscr{M}_1, \mathscr{M}_2, ..., \mathscr{M}_n\}$ where $\mathscr{M}_i$ is an individual model representing one possible exploration pattern. Depending on the choice model space, these individual models could inform us about user's bias in certain attributes of data, a specific task a user may be performing, or something about a user's personality. In this work, we encode exploration patterns in terms of which subset of data dimensions drive user exploration. For a dataset with d dimensions, we create a model space of 2^d models, where each model represents exploration based on a subset of d dimensions. As a user interacts with a visualization, we maintain a belief over viability of each model given the observed interactions. The belief over our models is incrementally updated using the Bayes' rule:

$$p(\mathscr{M}_i \mid \text{interactions}) \propto p(\text{interactions} \mid \mathscr{M}_i) p(\mathscr{M}_i), \quad (1)$$

where each model is evaluated in light of observed interactions and prior beliefs. Next, we discuss the details of initiating a prior belief over the model space and updating the belief as the user interacts with data points.

4.1 Maintaining a Belief over Competing Models

Before observing any interactions, our **prior** assumes a uniform distribution over the set of all possible models. That is, for every model, $\mathscr{M}_i$,

$$p(\mathscr{M}_i) = \frac{1}{2^d} \quad (2)$$

This choice of prior reflects the idea that we are uniformly uncertain at the beginning about what combination of attributes best describe exploration patterns. In other words, we consider every exploration pattern to be equally likely before observing any interactions. In Section 7.2, we will demonstrate an alternative prior that penalizes models which encompass exploration bias towards a large number of attributes.

As new interactions occur, we update these models throughout a session via the process of iterative Bayesian updating. That is, given the set of interactions, $\mathscr{C}$, the **posterior** belief over the set of models is:

$$p(\mathscr{M}_i \mid \mathscr{C}) \propto p(\mathscr{C} \mid \mathscr{M}_i) p(\mathscr{M}_i) \quad (3)$$

where $p(\mathscr{C} \mid \mathscr{M}_i)$ is the **likelihood** representing how well a model $\mathscr{M}_i$ can explain the observations in $\mathscr{C}$. We can expand the likelihood function via chain rule to get:

$$p(\mathscr{C} \mid \mathscr{M}_i) = \prod_{\vec{c}_j \in \mathscr{C}} p(\vec{c}_j \mid \mathscr{M}_i, \mathscr{C}_{1:j-1}) \quad (4)$$

In Eq. 4, $\vec{c}_j$ is the jth click and $\mathscr{C}_{1:j-1}$ is the set of first $j-1$ clicks. Furthermore, notice that $p(\vec{c}_j \mid \mathscr{M}_i, \mathscr{C}_{1:j-1})$ is a measure of how relevant $\vec{c}_j$ is given model $\mathscr{M}_i$ in light of all past observations.

This framework of competing models thus far is flexible to operate regardless of visualization design or dataset. Picking a likelihood function is the key part to fitting this framework into specific applications. For example, if our application involves a collection of text documents as the underlying dataset, we need to choose a likelihood function that describes user interactions with support over the set of text documents. If our application involves multidimensional data with continuous and discrete attributes, we need to choose a likelihood function that describes user interactions with support over all data points. In Section 4.2, we narrow the scope of possibilities and consider a specific likelihood function that models information relevance in multidimensional datasets with continuous and discrete attributes.

4.2 Model Definition for Multidimensional Data

The focus of this section is on modeling how relevant a data point is to the user given their past interactions. This measure of relevance is denoted in Eq. 4 as $p(\vec{c}_j \mid \mathscr{M}_i, \mathscr{C}_{1:j-1})$. In more formal terms, we want to use the evolving interaction data to infer the user's **data objective,** which we define to be some user-determined probability distribution $p(\vec{x})$ over the data space with support over the items of interest. Points

with higher values of $p(\vec{x})$ are considered more **relevant** to the session. Our framework requires the models to be probabilistic (i.e. include information about uncertainty in outputs). There are multiple techniques for inferring the relevance of a data point to the user in light of past observations. Some examples include probabilistic classification techniques such as logistic regression and k-nearest neighbors model. This freedom in model choices makes our approach design-agnostic. Depending on the scenario and the dataset, practitioners may encode exploration patterns into appropriate models. In this work, we model the distribution of data points with which the user has interacted. This proof-of-concept approach provides us with a simple and interpretable method to explore the idea of competing models. Since the multi-dimensional datasets in this paper involve discrete and continuous dimensions, we utilize two kinds of parametric distributions: Gaussian and categorical. While the parametric nature of these distributions impose some restrictions (e.g. unimodality of the Gassian), their parameters make the distributions easier to interpret by simply observing the value of parameters.

Continuous Dimensions   The Gaussian distribution is commonly used for continuous variables. We utilize this family of distributions to model user's interest in each continuous dimension of the data. For every model, $\mathcal{M}_i$, we construct a $|\mathbb{A}_{c,i}| + 1$ dimensional multivariate Gaussian to learn the distribution of observations for continuous attributes as well as the timestep (*time*) of interactions, where $|\mathbb{A}_{c,i}|$ is the number of continuous dimensions involved in a given model $\mathcal{M}_i$ and an extra dimension is added to model the order at which observations arrive. Multivariate Gaussian distribution is parameterized by a mean vector $\vec{\mu}$ and a covariance matrix Σ, representing center and spread of the distribution respectively. Since the value of these parameters $(\vec{\mu}, \Sigma)$ for a given set of observations $\mathcal{C}$ is unknown, we infer them thorough the Bayesian inference process which results in a closed form posterior predictive distribution function, $f_c(\vec{x} \mid \mathcal{C})$. This probability distribution function, f_c is known as Student's t-distribution [2, 31]. Moreover, the Bayesian inference for this distribution involves four hyper-parameters. We set these hyper-parameters so that the prior belief is *uninformative* (i.e. contains no specific information about user exploration before interactions are observed). The derivation details of this posterior predictive distribution as well as the choice of hyper-parameters are further discussed in the supplementary material. Finally, we normalize the probability density function f_c across all data points in $\mathcal{D}$ to assign a probability value to the continuous dimensions of every point $\vec{x} \in \mathcal{D}$:

$$p_c(\vec{x} \mid \mathcal{M}_i, \mathcal{C}_{1:t}, time = t+1) = \frac{f_c(\vec{x}[\mathbb{A}_{c,i}] \mid \mathcal{C}_{1:t}, time = t+1)}{\sum_{\vec{x}' \in \mathcal{D}} f_c(\vec{x}'[\mathbb{A}_{c,i}] \mid \mathcal{C}_{1:t}, time = t+1)} \tag{5}$$

In the equation above, $\vec{x}[\mathbb{A}_{c,i}]$ denotes the continuous dimensions of $\vec{x}$ that are relevant in model $\mathcal{M}_i$.

Discrete Dimensions   The categorical model is used to explain the probability of discrete events occurring. For an attribute domain with K possible categories, the categorical model has a K-dimensional vector $\vec{\mu}$ which describes the probability of observing each of the K choices. Since the value of $\vec{\mu}$ is unknown, we infer it via the Bayesian inference process. The Bayesian inference process for the categorical distribution relies on one hyper-parameter $(\vec{\alpha})$ representing the pseudocount for the prior. We set this hyper-parameter so that the prior is uninformative as demonstrated in Fig. 5 (i.e. contains no specific information about user exploration before interactions are observed). Let d' be a categorical attribute with K possible values. The posterior predictive of observing category k given a set of interactions $\mathcal{C}$ is [2, 40]:

$$f_{d'}(k \mid \mathcal{C}, \vec{\alpha}) = \frac{\alpha_k + m_k}{\sum_{i=1}^{K} (\alpha_i + m_i)} \tag{6}$$

where α_i is the pseudocount for category i (from hyper-parameter $\vec{\alpha}$) and m_i is the observation count for category i. We offer more details on this hyper-parameter and the derivation of closed-form posterior predictive in the supplemental material. We normalize the posterior

predictive of each categorical distribution across all data points in $\mathcal{D}$ to assign a probability value to each discrete dimension of every point $\vec{x} \in \mathcal{D}$. For a discrete dimension d' we have:

$$p_{d'}(\vec{x} \mid \mathcal{C}_{1:t}) = \frac{f_{d'}(\vec{x}[d'] \mid \mathcal{C}_{1:t})}{\sum_{\substack{\vec{x}' \in \mathcal{D} \\ \text{s.t. } \vec{x}[d'] = \vec{x}'[d']}} 1} \tag{7}$$

In the equation above, $\vec{x}[d']$ is the value of attribute d' for data point $\vec{x}$.

Combined Point Likelihood   Equations 5 and 7 provide us with probability distributions over continuous and discrete dimensions of the dataset respectively. The overall probability of any point $\vec{x} \in \mathcal{D}$ within a model $\mathcal{M}_i$ in light of past t interactions $\mathcal{C}_{1:t}$ is:

$$p(\vec{x} \mid \mathcal{M}_i, \mathcal{C}_{1:t}) = p_c(\vec{x} \mid \mathcal{M}_i, \mathcal{C}_{1:t}, time = t+1) \prod_{d \in \mathbb{A}_{d,i}} p_d(\vec{x} \mid \mathcal{C}_{1:t}) \tag{8}$$

This section provided a detailed explanation for initiating and maintaining a belief over a set of competing models. Specifically, section 4.1 discussed the *prior* and *posterior* portions of equation 3, and section 4.2 discussed the *likelihood* portion of equation 3. Going forward, we apply these concepts to detect exploration bias, predict future clicks, and summarize an analytic session.

5   APPLICATIONS OF COMPETING MODELS

The ultimate goal of our technique is to make high-level and in-depth inferences by observing their low-level user interactions. More specifically, our technique takes a more comprehensive view of user modeling by enabling us to *infer exploration bias*, *predict future interactions*, and *summarize an analysis session*. In this section, we use our competing models framework (Section 4) in order to gain insight into users.

5.1   Exploration Bias Detection

Bias often has a negative connotation; however, in reality bias can be desirable. In very simple cases, bias can indicate the criteria in a data space on which people make decisions [41]. Therefore, we categorize bias into two categories: intentional and unintentional. Intentional bias holds information about the users' search criteria and can accelerate the analysis process. Unintentional bias, on the other hand, arises from the user subconsciously avoiding some information and exploring others based on some personal factor. Unintentional bias leads to missing information. Regardless of the category, detecting biases in visualization systems can improve the quality of analysis and decision.

Throughout a session, our framework maintains a posterior belief over the set of models $\{\mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_{2^d}\}$ each of which represent biased exploration towards a subset of attributes. The normalized posterior belief, $\{p(\mathcal{M}_1 \mid \mathcal{C}), p(\mathcal{M}_2 \mid \mathcal{C}), ..., p(\mathcal{M}_{2^d} \mid \mathcal{C})\}$, informs us about which subset of attributes the user may be biased towards. Higher values of $p(\mathcal{M}_i \mid \mathcal{C})$ indicate higher chances of the user being biased towards attributes represented in $\mathcal{M}_i$.

Using the normalized posterior belief over models, $\{p(\mathcal{M}_1 \mid \mathcal{C}), p(\mathcal{M}_2 \mid \mathcal{C}), ..., p(\mathcal{M}_{2^d} \mid \mathcal{C})\}$, we can use the law of total probability to calculate the chance of bias towards each individual attribute:

$$p(\text{bias towards attribute } a \mid \mathcal{C}) = \sum_{\mathcal{M}_i \in \mathbb{M}_a} p(\mathcal{M}_i \mid \mathcal{C}) \tag{9}$$

where $\mathbb{M}_a$ denotes the subset of models in which bias towards attribute a is assumed.

5.2   Next Interaction Prediction

Predicting which data points the user may interact with next opens opportunities to add features such as target assistance and target gravity to help users find the next most interesting data points in their exploration session [33]. Anticipating next clicks given past clicks involves a time component. Adding time-steps as an additional dimension to our continuous model enables us to correlate clicks not only by their attributes, but also by the time-step at which they were clicked. For

each of the data points, $x \in \mathcal{D}$, we calculate the probability of the user interacting with x at the next time-step via:

$$p(\vec{x} \mid \mathcal{C}_{1:t}, time = t+1) =$$
$$\sum_{\mathcal{M}_i \in \mathbb{M}} p(\vec{x} \mid \mathcal{M}_i, \mathcal{C}_{1:t}, time = t+1) p(\mathcal{M}_i \mid \mathcal{C}_{1:t}) \quad (10)$$

The Bayesian approach for not losing any information by selecting one single model is *Bayesian model averaging*, which we use in the equation above.

5.3 Summarizing Analytic Session

Using provenance data to summarize analytic sessions has been of interest in the visualization community [17, 20, 44]. Due to our parametric choice of models in Section 4.2, we show that those parameters can be used to represent a summary distribution from which the clicks were generated. For example, let a be a continuous attribute. The updated parameters of the Gaussian distribution (μ_a, σ_a^2) after observing interactions can display a summary of what distribution of the values in attribute a the user has explored. In the discrete case, the parameter of categorical distribution serves as a summary of observed interactions in a certain categorical domain. Section 7 demonstrates this effect through figures.

6 END-TO-END EXAMPLE

In this section, we consider an end-to-end example with a small dataset to demonstrate our framework from Sections 4 and 5. Let $\mathcal{D}$ be a set of seven fictitious restaurants with *location* and *food type* being continuous and discrete attributes respectively. We encode each restaurant into a 3D vector of form $(latitude, longitude, type)$ to get:

$$\mathcal{D} = \{(0.35, 0.85, \text{Italian}), (0.8, 0.35, \text{Mexican}), (0.85, 0.1, \text{Persian}),$$
$$(0.7, 0.3, \text{Italian}), (0.15, 0.75, \text{Mexican}), (0.1, 0.05, \text{Persian}),$$
$$(0.9, 0.85, \text{Mexican})\}$$

We assume that exploration based on latitude or longitude alone is likely unnatural, and combine the two to be a single 2D continuous attribute, *location*. Hence, the set of continuous attributes is $\mathbb{A}_c = \{location\}$ and the set of discrete attributes is $\mathbb{A}_d = \{type\}$. Since the user may explore this dataset based on any combination of these attributes, we construct $2^{|\mathbb{A}_c| + |\mathbb{A}_d|} = 4$ models to represent exploration patterns (Table 2). After two timesteps, we observe that the user has interacted with the following set of points (in the order listed):

$$\mathcal{C} = \{(0.85, 0.1, \text{Persian}), (0.8, 0.35, \text{Mexican})\}$$

The interaction type here is irrelevant, but for the sake of example we could assume it occurs through clicking, hovering, or adding to favorites.

Table 2. Example model space $\mathbb{M}$, where each model represents exploration based on a subset of attributes. ✔ indicates an attribute is considered important to the exploration session and ✘ indicates an attribute is not considered important to the exploration session.

Model	loc.	type	$p(\mathcal{M}_i)$	$p(\mathcal{M}_i \mid \mathcal{C}_{0:1})$	$p(\mathcal{M}_i \mid \mathcal{C}_{0:2})$
$\mathcal{M}_1$	✘	✘	0.25	0.08	0.03
$\mathcal{M}_2$	✔	✘	0.25	0.12	0.10
$\mathcal{M}_3$	✘	✔	0.25	0.29	0.16
$\mathcal{M}_4$	✔	✔	0.25	0.51	0.70

Before any interactions are observed, we are uniformly uncertain about which combination of attributes drive interactions ($p(\mathcal{M}_i)$ column in Table 2). As more clicks arrive, our posterior belief gains confidence that *location* and *type* attributes explain the interactions ($p(\mathcal{M}_4 \mid \mathcal{C}_{0:1}) = 0.51$ and $p(\mathcal{M}_4 \mid \mathcal{C}_{0:2})$ in Table 2). The posteriors

(columns $p(\mathcal{M}_i \mid \mathcal{C}_{0:1})$ and $p(\mathcal{M}_i \mid \mathcal{C}_{0:2})$) are computed using a normalized probability mass function (Figure 1) as the likelihood function.

In order to *detect exploration bias* towards a certain attribute, we use Eq. 9 to compute the marginal probability of models involving that attribute. For example, the probability of bias towards location after observing two clicks is 0.80, since $p(\mathcal{M}_2 \mid \mathcal{C}_{1:2}) + p(\mathcal{M}_4 \mid \mathcal{C}_{1:2}) = 0.8$ and models $\{\mathcal{M}_2, \mathcal{M}_4\}$ both involve the *location* attribute.

For *next click prediction*, we compute the likelihood of every every point averaged over all models. For every point $x \in \mathcal{D}$, we compute:

$$\sum_{\mathcal{M}_i \in \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4\}} p(x \mid \mathcal{M}_i, \mathcal{C}_{1:2}, time = 3) p(\mathcal{M}_i \mid \mathcal{C}_{1:2})$$

which results to $(\text{Mexican}, 0.15, 0.75)$ being the predicted next interaction (assuming no repeated clicks). More specifically, the order of points for being the next interaction is as follows:

$$(\text{Mexican}, 0.8, 0.35) > (\text{Persian}, 0.85, 0.1) > (\text{Mexican}, 0.15, 0.75) >$$
$$(\text{Mexican}, 0.9, 0.85) > (\text{Persian}, 0.1, 0.05) > (\text{Italian}, 0.7, 0.3) >$$
$$(\text{Italian}, 0.35, 0.85)$$

Notice that *location* and *type* driving exploration (as inferred in Table 2) is reflected in the ordering above where the top choices for a third click (assuming no repeated clicks) are Mexican and Persian restaurants in order of their proximity to past interactions. Moreover, notice that the distributions in Figure 1 start with uninformative priors when $t = 0$, but update to *summarize interactions* as they arrive ($t = 1$ and $t = 2$).
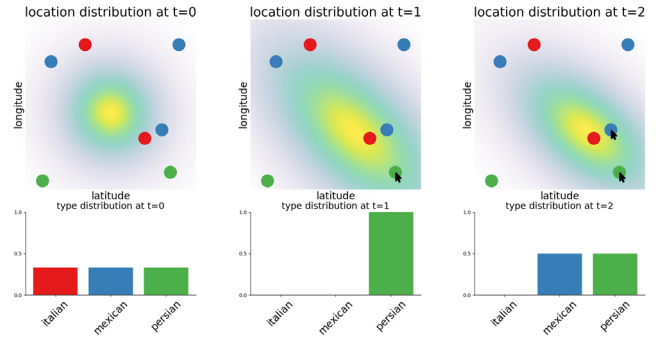

Fig. 1. The inferred Gaussian and categorical distributions *summarizing the first two interactions*. As clicks (denoted by ➤) arrive, our models get updated to represent user interactions.

7 VALIDATION WITH USER STUDY DATASETS

We validate our proposed technique on three user study datasets. Each dataset seeks to highlight a unique aspect of our technique. First, we validate our technique using a user study dataset collected by Ottley et al. [33] where users are asked to perform specific tasks using a map visualization of crimes in St. Louis. Since this user study dataset includes ground truths associated with each session, we can compare our outcomes of *exploration bias detection* and *next interaction prediction* with established baselines by Wall et al. [41] and Ottley et al. [33] respectively.

Second, we demonstrate our technique's independence from visualization design by validating it on a user study dataset collected by Feng et al. [15] where users freely interact with a visualization of S&P 500 companies. While this dataset still involves a point-based visualization, it differs from the first dataset in that it is not a map-based visualization and the interaction technique is *hovers* instead of *clicks*. Considering that the ground-truth bias in these open-ended sessions are unknown, we are only able to compare the outcome of *next interaction prediction* with the established baseline by Ottley et al. [33].

Lastly, we demonstrating our technique's independence from tasks by validating it on an open-ended user interaction session with the map of crimes in St. Louis, and compare our models' understanding of user exploration to the self-reported insight by the participant.
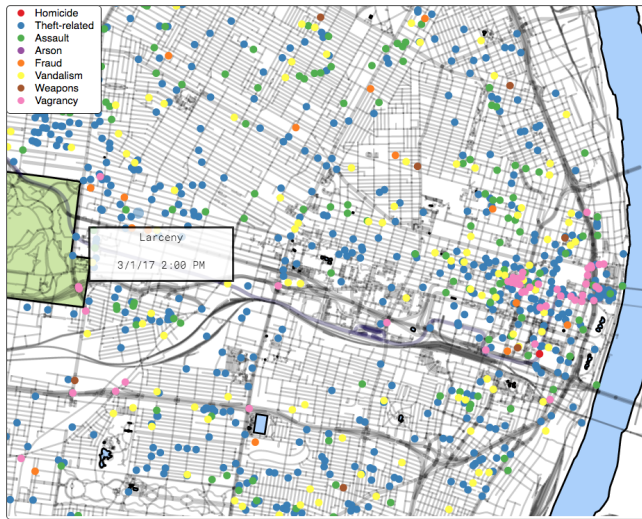
Fig. 2. The map of St. Louis crimes used by Ottley et al. [33]. The position of the dots indicate the *location* of the crime, and the color of the dots indicate the *type* of crime. Users explored the data by clicking on the dots and observing more information in a tooltip.

Table 3. Number of participants and ground truth exploration biases for each category of tasks in the Ottley et al. experiment [33]

Task Category	Ground-Truth Bias	# of Participants
Geo-Based	*latitude, longitude*	28
Type-Based	*type*	23
Mixed	*latitude, longitude, type*	27

7.1 Map of Crimes in St. Louis

7.1.1 Ottley *et al.* experimental setup

Ottley et al. [33] describe a user study in which they capture mouse click data as participants interact with a map visualization of crimes in Saint Louis. In this experiment, participants were presented a set of 1,951 reported crimes from March 2017 visualized on an interactive map. Each instance was displayed as a dot with a position and a color indicating the *location* and *type* of the crime respectively. The map responded to user clicks by triggering a tooltip containing more details on the crime. The questions presented to the participants required them to search the map to find an answer based on the data:

1. Out of all the cases of Homicide, one case differs from the other cases with regard to time. What is the time of that case?
2. How many cases of arson occur during PM?
3. There are four types Theft-Related crime in the red shaded region: Larceny, Burglary, Robbery and Motor Vehicle Theft. Count the number of cases of Robbery in the red shaded region.
4. There are two types of Assault: Aggravated and Non-Aggravated assault. Count the number of Non-Aggravated Assault in the red shaded region
5. Count the number of crimes that occur during 7:00 AM - 12:30 PM in the red shaded region.
6. Count the number of crimes during AM in the red shaded region. The six questions above were categorized into three groups of tasks: *geo-based* (search specific location, questions 5-6), *type-based* (search for specific type of crime, questions 1-2), and *mixed* (search for a particular type of crime in a specific region of the map, questions 3-4). Table 3 summarizes the number of participants for each group of tasks.

7.1.2 Results

In our first set of analysis, we consider ***exploration bias detection*** based on the ground-truths presented in Table 3, and we compare the output of our technique from Eq. 9 with the equivalent baseline metric. As discussed in Section 2, Wall et al. [41] propose an *attribute distribution* metric to measure how biased an exploration session is towards any

particular attribute. They define this metric to be the complement of observing the set of interactions assuming the distribution of attributes within the full dataset and interaction dataset are the same. They specifically suggest the non-parametric KS test for continuous attributes and Chi-Square test for discrete attributes. For each attribute, these tests take two sets of 1-dimensional sets as inputs (full data and interaction data), and decide if the two sets are from the same distribution. Let p_a be the output of a KS or Chi-Square test for attribute a; then, the *attribute distribution* metric is defined as $b_{Ad}(a) = 1 - p_a$. Note that higher values of $b_{Ad}(a)$ correspond to more exploration bias towards attribute a. For tasks with bias towards more than one attribute, we compute the product of this *attribute distribution* metric to represent the conjunction of biases. A comparison between our algorithm's bias detection and the baseline is shown in Figure 3. After 12 clicks, our method significantly outperformed Wall et al. on the geo-based and mixed tasks according to a paired t-test with p-value 1.5×10^{-7} and 0.003, respectively. Wall et al. outperformed our method on the type tasks with p-value 0.1. Upon further investigation, we concluded that the better performance of baseline in type-based sessions is explained by the characteristic of corresponding type-based questions. Questions 1 and 2 ask participants to explore under-represented categories of the data (i.e. Homicide and Arson), and Chi-Square tests can detect that bias quicker. On the other hand, our technique outperforms the Chi-Square baseline when detecting bias towards over-represented categories. While the user study datasets did not include sessions to reflect this effect, we verified them using synthetic sessions from the crime dataset.

In our second set of analysis for this dataset, we consider ***next click prediction*** and compare the outcome of our technique with the baseline from Ottley et al. [33]. This baseline is based on the idea that users click on items that are within a close proximity from each other. The implementation of this baseline involves iteratively sampling particles from a current belief, using the observation model to re-weight the particles based on observed clicks, and ranking the datapoints according to the weight and proximity of particles at each timestep. Then, the set of top-k candidates are selected to represent a prediction set for the next datapoint with which the user may interact. In order to be consistent with the existing work, we follow a similar approach in our technique: after averaging our competing models to compute the predictive posterior of each data point being the next observation as outlined in Eq. 10, we sort them based on ranking and pick the top-k data points as the set of candidates for next interaction. At each time-step, we determine if our prediction set included the next click and compute the rate of success in this prediction process. A comparison between our algorithm's next click prediction and the baseline is shown in Figure 4. Across all three tasks, our technique outperforms this baseline for the majority of k values and performs within the margin of error for the rest. Anticipating future data points with which the user may interact with has implications in designing more intelligent visualization systems which understand user goals and assist them in finding the information they desire. We discuss more of this topic in Section 9.

The final goal of our technique is to ***summarize analytic sessions***. Since we chose parametric models to represent user interactions, summarizing a session is just a matter of observing the value of parameters at each time-step. Figure 5 shows the progression of our algorithm from as new interactions arrive. By only looking at the summary of interactions through time, we can easily map the session to question 4, where the user is exploring Assault cases in a particular region of St. Louis.

7.2 S&P 500 Boardrooms

Thus far in our validation, we have relied on a map visualization with which users have to interact through clicks in order to accomplish pre-defined tasks. In this part of our validation, we consider a multi-section visualization of S&P 500 board of directors published by Wall Street Journal [27]. The underlying data for this visualization has one discrete and six continuous attributes. A tooltip containing these attributes appears on hover. A view of this visualization is shown in Figure 6.
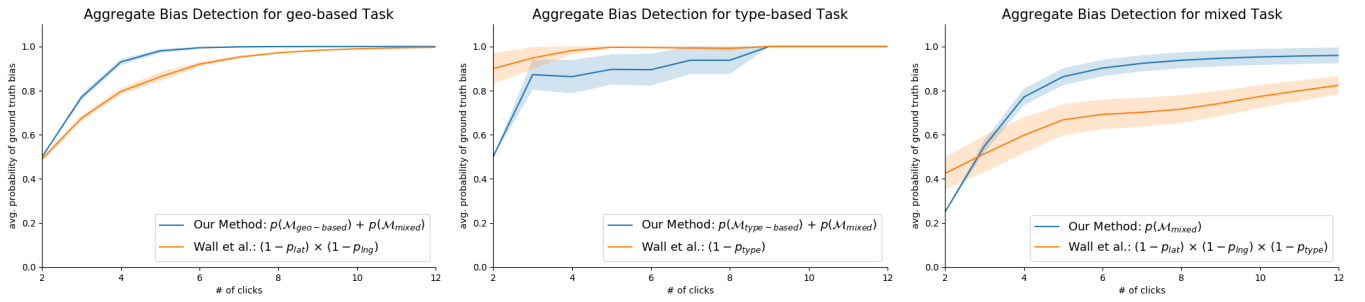
Fig. 3. Results of **exploration bias detection** on user study dataset from the Ottley et al. experiment (Section 7.1). The participants were asked to interact with a map visualization of crimes in St. Louis in order to answer geo-based questions *(based on latitude and longitude)*, type-based questions *(based on type)*, and mixed questions *(based on latitude, longitude, and type)*. After 12 clicks, our method significantly outperformed Wall et al. on the geo-based and mixed tasks according to a paired t-test with p-value 1.5×10^{-7} and 0.003, respectively. Wall et al. outperformed our method on the type tasks with p-value 0.1. The left graph shows that our method detects exploration bias towards *location* quicker and more confidently than the baseline. The middle graph shows that our technique detects exploration bias towards the *type*, however, it takes longer than the baseline to gain confidence in the level of bias (discussed in Section 7.1.2). The right graph shows that our method detects bias towards *location and type* quicker and more confidently than the baseline. The shaded region in all three graphs represent the standard error among 28 geo-based sessions, 23 type-based sessions, and 27 mixed sessions.
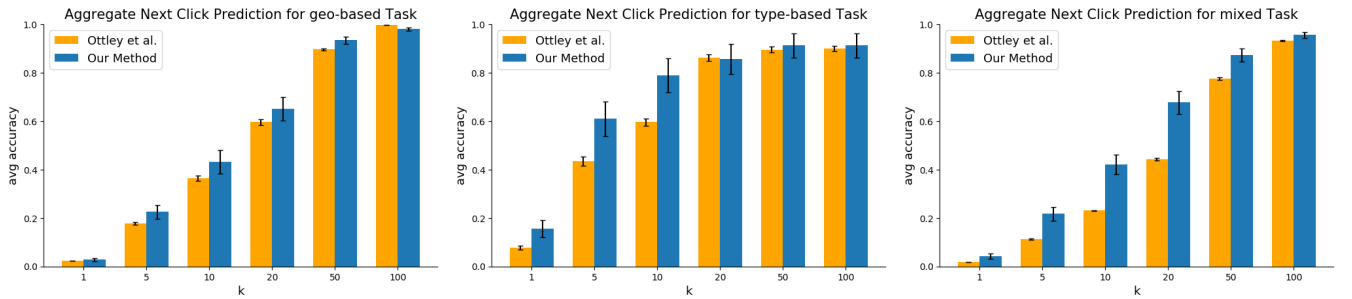

Fig. 4. Results of **next click prediction** on user study dataset from the Ottley et al. experiment (Section 7.1). Our method performs within the margin of error of the previously studied hidden Markov model in all categories, while outperforming it in most cases. The error bars in all three graphs represent the standard error among 28 geo-based sessions, 23 type-based sessions, and 27 mixed sessions.
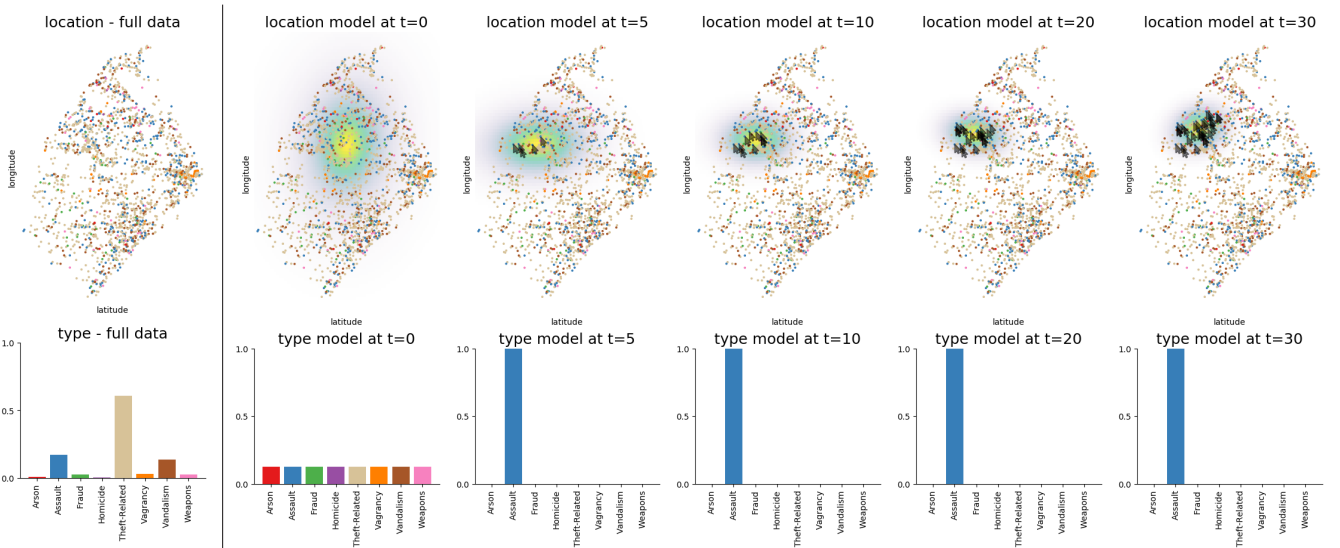

Fig. 5. An instance of **summarizing analytic sessions**, where the user is performing mixed interactions *(based on location and type)* in order to answer question 4 from Section 7.1. The left-most graph shows the distribution of the full crime dataset presented to participants. At $t = 0$ before any clicks have been observed, our models reflect our prior. As clicks (denoted by) arrive, our models get updated to represent user interactions. Notice the significant difference in *type model* from $t = 0$ to $t = 5$. This is because our pseudo-count for the prior is small, making the prior insignificant in comparison to observed user clicks.
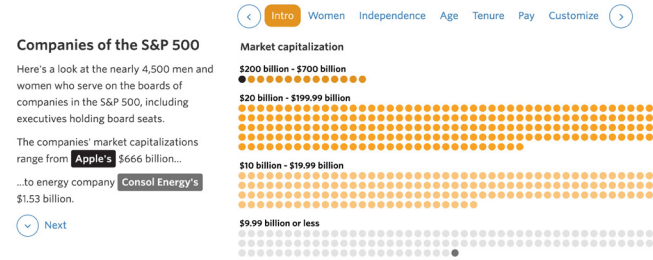
Fig. 6. "Inside America's Boardrooms" a multi-section visualization published in the Wall Street Journal presenting data on leaders of S&P 500 companies [27]. Users from the Feng et al. experiment [15] interacted with this visualization by *hovering* over different companies and observing more information about the leaders of the companies in a tooltip.

7.2.1 Feng *et al.* experimental setup

Feng et al. [15] collected user interaction sessions, where users freely interact with point-based elements in a visualization and search for keywords. Specifically, they use the visualization called "Inside America's Boardroom" published in the Wall Street Journal (Fig. 6). This underlying data for this visualization contains 7 dimensions: *market capitalization, ratio of unrelated board members, ratio of female board members, average age of board members, average tenure, median pay, and industry group*. As opposed to the Ottley et al. experiment, participants in this study were asked to perform an open-ended exploration on the dataset. The recorded sessions contain a sequence hovers on different data points and the duration of each hover. While this user study was designed to study the impact of having search capabilities in visualizations, we utilize the hover data to validate our technique for predicting future hovers. Since hovers can inherently be noisier than clicks due to unintentional hovers while going from one data point to another, we filtered the sessions to only include hovers that lasted for over one second. This will eliminate most unintentional hovers that occurred in transition of going from one data point to another. Moreover, we filtered the data further to include only those with more than 3 hovers. Before filtering, there were 41 sessions with avg. 391.19 hovers per sessions (SD 237.75). After filtering, we had 39 sessions with avg. 26.83 hovers (SD 18.95) and at least 3 hovers per session.

7.2.2 Results

In our analysis of this user study, we consider *next hover prediction*. We modify the Ottley et al. baseline from Section 7.1 to include all seven attributes of this dataset. Our technique creates a space of 2^7 models (corresponding to all possible subsets of the 7 dimensions), where each model represents biased exploration towards a certain subset of attributes. Furthermore, we make an adjustment to the prior belief over the model space in order to penalize models that encompass bias towards a large subset of attributes. In Section 4.1, we suggested the uniform prior belief $p(\mathscr{M}_i) = 1/2^d$, however, here we use $p(\mathscr{M}_i) \propto 1/(d+1)$, where d'_i is the number of attributes represented in model $\mathscr{M}_i$. Figure 7 shows that our technique outperforms the established baseline for higher values of k and performs within the margin of error for lower k.

7.3 Open-ended tasks with Map of Crimes in St. Louis

In our final phase of validation, we reconsider the map of crimes in St. Louis. The experiments in Section 7.1 involved narrowly defined tasks. In this case-study, we have selected one analytic session where the user freely interacted with the visualization and self-reported an insight they had. The main purpose of this final case study is to highlight how the self-reported insight relates to our model's summary of session.

7.3.1 Kern *et al.* experimental setup

In this crowdsourced experiment, Kern et al. [25] propose an alternative experimental design in which participants freely interact with a map
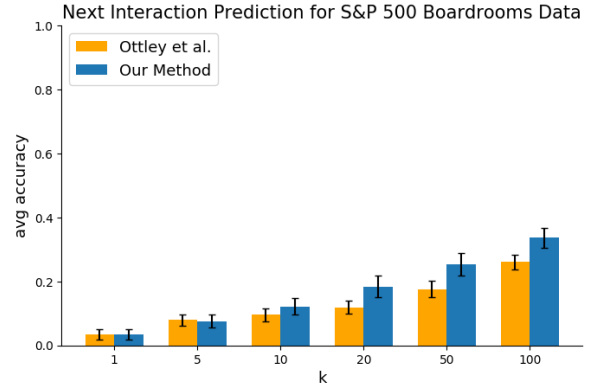

Fig. 7. Average accuracy of predicting the next hover in the set of top-k candidates for boardrooms dataset. There are 39 sessions, and prediction starts after the third click for consistency with the baseline from Ottley et al. [33]. The error bars indicate the standard error of accuracy across all sessions. There are 500 data points in the visualization, making $k = 50$ to be 10% of the dataset. Our technique outperforms or performs within margin of error in comparison to the baseline.
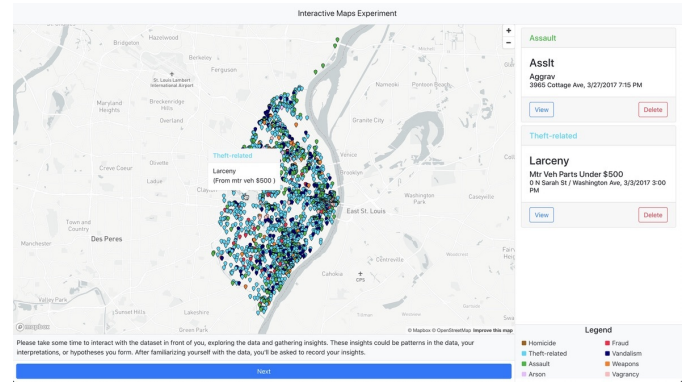

Fig. 8. The experiment interface from Kern et al. [25]. Crime cases were visualized on an interactive map, where the colors indicate the type of crimes. Users were able to zoom, pan, hover, and click. While hovering on dots, a tooltip with more details opens. When a dot is clicked, the crime case is added to the sidebar as shown above.

visualization of crimes in St. Louis and report their insights. Figure 8 shows the visual interface for this experiment, hovers trigger a tooltip with more details and clicks add the data point to the side bar. In this paper, we used clicks from one session that had ≥ 10 click events. Before the participants start the experiment, they are asked to familiarize themselves with the interface and functionalities. We use recorded clicks from this experiment to study how our model's understanding towards the user compares with the self-reported insight.

7.3.2 Results

After observing six clicks in this session, our model posterior inferred the user is exploring based on only location (geo-based) with probability > 0.8. By the time eight clicks were observed, our model was more certain on this task being geo-based with probability > 0.95. Notice that this finding is trivially justified in Figure 9, where the distribution of *type* attribute in the interaction set is very similar to the full data distribution (hence no bias towards type), but the distribution of *location* attribute in the interaction set is narrow compared to the full dataset. Our model's understanding of the user reflects their self-reported insight: *"Two cars were stolen in the same week from the 3900 block of Miami St."* We hypothesize that the absence of bias towards *type* is due to the user looking for an insight among different *type*s of crimes in the session.
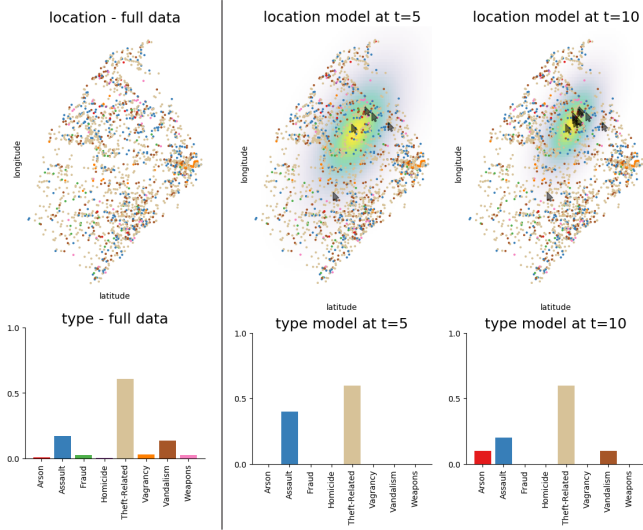
Fig. 9. The summary of a session in which the user is freely interacting with a map visualization to report their insights. The left-most graph shows the distribution of the full crime dataset presented to the participant. As clicks (denoted by ⬏) arrive, our models get updated to represent user interactions. Notice that type model at time $t = 10$ does not significantly differ from the type distribution in the full data set, resulting in no detection of exploration bias towards *type*.

8   Discussion

The novelty of our technique is that we encode different exploration patterns as models and use the Bayesian model selection to maintain a belief over all possible exploration patterns. To demonstrate how a set of competing models can uncover user exploration patterns, we designed our model space so that each model represents information relevance based on a subset of dimensions in data.

One of the immediate insights from our results is that posing user exploration modeling as a Bayesian model selection problem results in better performance in *exploration bias detection* and *next interaction prediction* in most cases. In particular, our choice of likelihood function from Section 4.2 resulted in improved performance for sessions in which the users were asked to explore a specific subset of the data (Figures 3, 4, and 5). In the case of bias detection for type-based tasks (Fig. 3, middle), the baseline outperformed our technique. After further investigation, we concluded that this behavior was due to the Wall et al. [41] baseline taking into account the overall data distribution whereas our likelihood function does not take into account the overall data distribution. An alternative choice of likelihood function that takes the underlying data distributions into account may result in improvements.

For open-ended tasks, we believe there could be more expressive choices of models to learn multi-modal distributions or more complex decision boundaries. Although our technique outperformed the existing baseline for open-ended tasks (Figure 7), we believe a more expressive choice of models and a more specific user study to prevent users from unintentional hovers would be beneficial. Another limitation of our work is that interactions that manipulate the underlying dataset are not supported as we assume our dataset $\mathscr{D}$ is constant throughout the session. In other words, we could consider interactions such as drag/drop to learn which subset of the data one may interact with, however, further investigation is needed to build models that learn about the context of re-positioning and make more complex predictions (such as where should a given point move given past movements).

The technique presented in this paper takes the visualization community a step closer to modeling users in scenarios where there are multiple possible exploration strategies, analysis goals, or personal preferences. By representing each possibility as a model and updating the belief as evidence is observed (through interactions), we can enable intelligent machine response in visualization systems. However, further work is required to address some of the shortcomings of our study.

We have identified three main areas for further exploration, which we discuss in the next section.

9   Future Directions

Framing user exploration modeling as a Bayesian model selection problem opens a new set of research opportunities. In particular, we can utilize recent advancements in Bayesian machine learning (i.e. active learning, active search, etc.) to further improve the collaboration between human and computer in visual analytics systems. In this section, we highlight four specific areas for future investigations.

Different Model Spaces   Visual analytics research is constantly evolving to study new human-related factors that impact how we use interactive visualization systems. This evolution in user modeling alongside our competing models framework provide an opportunity to infer which human-related factor is impacting an individual session the most and move towards more individualized visual analytics systems.

Sampling Methods for More Efficiency   Throughout this work, we compute exact values of model posteriors. While this was not prohibitive for our largest model space (2^7 models in Section 7.2, where there were 7 attributes), we recognize computing exact posteriors is not always feasible. This could be due to the unavailability of posterior predictive distributions in closed form or the need for low latency in real-time systems. In such scenarios, sampling methods may be used to approximate the posterior distribution.

Active Learning to Mitigate Information Bubble   Active learning is the idea behind an algorithm that queries an oracle in order to learn. In the scope of our work, this concept can be utilized to confirm detected biases with the user in order to avoid unintentional biases (through *exploration*) and mitigate information overload by focusing on relevant information (through *exploitation*). The mechanics of effectively querying users and updating the visualization view are also left for future studies. This line of research can lead to building *automatic filtering* features to mitigate information bubbles created by unintentional bias.

10   Conclusion

We began this paper by claiming that interactive visualizations alone cannot fully support humans in analyzing large datasets and making informed decisions. There are numerous human factors such as limited cognitive capacity and personal biases that have roles in hindering effective analysis of data. As the visual analytics research community investigates more about human factors related to data analysis, we need to build intelligent visual analytic systems who understand users through the lens of their low-level interactions and ultimately improve the visual analytic experience. In order to enable machine teammates to make high-level inferences by observing low-level interactions, we proposed *competing models*: a technique which enumerates a set of human-related possibilities as models and then utilizes iterative Bayesian updating to learn about users as interactions are observed. To narrow our focus, we outlined the process of creating and maintaining a set of competing models, and described how visual analytics researchers can use this idea to gain a deeper insight into user exploration. In the process of validating our technique, we demonstrated that by building models on the underlying dataset in visualizations and then updating said models as interactions are observed, we can uncover exploration biases and predict future interaction. In particular, we reached high rates of accuracy for *next interaction prediction* when the participants were asked to perform specific tasks. In open ended scenarios, we saw a lower overall performance while still outperforming the baseline. This observation calls for more investigation on modeling open-ended sessions, which is inherently a more difficult objective.

REFERENCES

[1] L. Battle, R. Chang, and M. Stonebraker. Dynamic prefetching of data tiles for interactive visualization. In *Proceedings of the 2016 International Conference on Management of Data*, pages 1363–1375. ACM, 2016.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

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

[4] Á. A. Cabrera, W. Epperson, F. Hohman, M. Kahng, J. Morgenstern, and D. H. Chau. Fairvis: Visual analytics for discovering intersectional bias in machine learning. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 46–56. IEEE, 2019.

[5] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo. Vistrails: visualization meets data management. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 745–747. ACM, 2006.

[6] Y. Chen, S. Barlowe, and J. Yang. Click2annotate: Automated insight externalization with rich semantics. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pages 155–162. IEEE, 2010.

[7] F. Dabek and J. J. Caban. A grammar-based approach for modeling user interactions and generating suggestions during the data exploration process. *IEEE transactions on visualization and computer graphics*, 23(1):41–50, 2016.

[8] N. Del Rio and P. P. Da Silva. Probe-it! visualization support for provenance. In *International Symposium on Visual Computing*, pages 732–741. Springer, 2007.

[9] M. Derthick and S. F. Roth. Enhancing data exploration with a branching history of user operations. *Knowledge-Based Systems*, 14(1-2):65–74, 2001.

[10] E. Dimara, A. Bezerianos, and P. Dragicevic. The attraction effect in information visualization. *IEEE transactions on visualization and computer graphics*, 23(1):471–480, 2016.

[11] J. Doboš, N. J. Mitra, and A. Steed. 3d timeline: Reverse engineering of a part-based provenance from consecutive 3d models. *Computer Graphics Forum*, 33(2):135–144, 2014.

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

[13] C. Dunne, N. Henry Riche, B. Lee, R. Metoyer, and G. Robertson. Graphtrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1663–1672. ACM, 2012.

[14] T. Ellkvist, D. Koop, E. W. Anderson, J. Freire, and C. Silva. Using provenance to support real-time collaborative design of workflows. In *International Provenance and Annotation Workshop*, pages 266–279. Springer, 2008.

[15] M. Feng, E. Peck, and L. Harrison. Patterns and pace: Quantifying diverse exploration behavior with visualizations on the web. *IEEE transactions on visualization and computer graphics*, 25(1):501–511, 2018.

[16] D. Gotz and M. X. Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, 2009.

[17] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit. From visual exploration to storytelling and back again. *Computer Graphics Forum*, 35(3):491–500, 2016.

[18] D. P. Groth and K. Streefkerk. Provenance and annotation for visual exploration systems. *IEEE transactions on visualization and computer graphics*, 12(6):1500–1510, 2006.

[19] G. S. Halford, R. Baker, J. E. McCredden, and J. D. Bain. How many variables can humans process? *Psychological science*, 16(1):70–76, 2005.

[20] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE transactions on visualization and computer graphics*, 14(6):1189–1196, 2008.

[21] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Queue*, 10(2):30, 2012.

[22] Z. Hensley, J. Sanyal, and J. New. Provenance in sensor data management: A cohesive, independent solution. *Communications of the ACM*, 57(2), 2014.

[23] T. Iwata, N. Houlsby, and Z. Ghahramani. Active learning for interactive visualization. In *Artificial Intelligence and Statistics*, pages 342–350, 2013.

[24] N. Kadivar, V. Chen, D. Dunsmuir, E. Lee, C. Qian, J. Dill, C. Shaw, and R. Woodbury. Capturing and supporting the analysis process. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pages 131–138. Ieee, 2009.

[25] A. Kern. The effects of mixed-initiative visualization systems on exploratory data analysis. https://openscholarship.wustl.edu/eng_etds/523/, 2020.

[26] D. Kurlander and S. Feiner. Editable graphical histories. In *VL*, pages 127–134. Citeseer, 1988.

[27] R. Lightner and T. Francis. Inside america's boardrooms. http://graphics.wsj.com/boards-of-directors-at-SP-500-companies/, 2016.

[28] H. R. Lipford, F. Stukes, W. Dou, M. E. Hawkins, and R. Chang. Helping users recall their reasoning process. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pages 187–194. IEEE, 2010.

[29] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen. Taxonomy-based glyph design—with a case study on visualizing workflows of biological experiments. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2603–2612, 2012.

[30] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE transactions on visualization and computer graphics*, 20(12):1633–1642, 2014.

[31] K. P. Murphy. Conjugate Bayesian analysis of the Gaussian distribution. *Technical Report*, 2007.

[32] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic provenance: process+ interaction+ insight. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 33–36. ACM, 2011.

[33] A. Ottley, R. Garnett, and R. Wan. Follow the clicks: Learning and anticipating mouse interactions during exploratory data analysis. *Computer Graphics Forum*, 38(3):41–52, 2019.

[34] S. G. Parker and C. R. Johnson. Scirun: a scientific programming environment for computational steering. In *Supercomputing'95: Proceedings of the 1995 ACM/IEEE Conference on Supercomputing*, pages 52–52. IEEE, 1995.

[35] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen. Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes. *IEEE transactions on visualization and computer graphics*, 22(1):31–40, 2015.

[36] W. Ribarsky and B. Fisher. The human-computer system: Towards an operational model for problem solving. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 1446–1455. IEEE, 2016.

[37] A. Sarvghad, M. Tory, and N. Mahyar. Visualizing dimension coverage to support exploratory analysis. *IEEE transactions on visualization and computer graphics*, 23(1):21–30, 2016.

[38] Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1237–1246. ACM, 2008.

[39] Y. L. Simmhan, B. Plale, D. Gannon, and S. Marru. Performance evaluation of the karma provenance framework for scientific workflows. In *International Provenance and Annotation Workshop*, pages 222–236. Springer, 2006.

[40] S. Tu. The Dirichlet-Multinomial and Dirichlet-Categorical models for Bayesian inference. *Computer Science Division, UC Berkeley*, 2014.

[41] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 104–115. IEEE, 2017.

[42] E. Wall, J. Stasko, and A. Endert. Toward a design space for mitigating cognitive bias in vis. In *2019 IEEE Visualization Conference (VIS)*, pages 111–115. IEEE, 2019.

[43] K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch. Survey on the analysis of user interactions and visualization provenance. *OSF Preprints*, 2020.

[44] S. Xu, C. Bryan, J. K. Li, J. Zhao, and K.-L. Ma. Chart constellations: Effective chart summarization for collaborative and multi-user analyses. *Computer Graphics Forum*, 37(3):75–86, 2018.