Guided Data Discovery in Interactive Visualizations via Active Search

Shayan Monadjemi*
Washington University in St. Louis

Sunwoo Ha†
Washington University in St. Louis

Quan Nguyen
Washington University in St. Louis

Henry Chai
Carnegie Mellon University

Roman Garnett
Washington University in St. Louis

Alvitta Ottley‡
Washington University in St. Louis

ABSTRACT

Recent advances in visual analytics have enabled us to learn from user interactions and uncover analytic goals. These innovations set the foundation for actively guiding users during data exploration. Providing such guidance will become more critical as datasets grow in size and complexity, precluding exhaustive investigation. Meanwhile, the machine learning community also struggles with datasets growing in size and complexity, precluding exhaustive labeling. Active learning is a broad family of algorithms developed for actively guiding models during training. We will consider the intersection of these analogous research thrusts. First, we discuss the nuances of matching the choice of an active learning algorithm to the task at hand. This is critical for performance, a fact we demonstrate in a simulation study. We then present results of a user study for the particular task of data discovery guided by an active learning algorithm specifically designed for this task.

Index Terms: Human-centered computing—Visual analytics; Human-centered computing—Empirical studies in visualization Computing methodologies—Active learning settings

1 INTRODUCTION

The visual analytics community has made significant strides in developing systems that facilitate the interplay between humans and machines in exploratory data analysis and sensemaking [9, 10, 13, 21, 23, 24, 35]. These advances have enabled us to learn from user interactions and uncover their analytic goals. Moreover, they have set up the foundation for creating visual analytic systems that guide users during data exploration. Providing such guidance will likely become more critical as datasets grow in size and complexity, overwhelming the limited screen real estate and human cognitive power.

The nuances of providing guidance varies with respect to factors including the visual analytic task at hand [7]. Prior work has explored a set of visual analytic tasks and corresponding techniques for providing guidance. For example Lin et al. [27] proposed a visual analytic system to aid users in detecting anomalous clusters in data. Similarly, Kucher et al. [25] presented a visual analytic system to aid users in building accurate stance classification models. Extending upon prior work, we focus on the common, yet tedious, task of discovery: sifting through a dataset and identifying valuable data points efficiently. For example, an intelligence analyst may spend substantial time reviewing documents while unraveling a terrorist attack plot, many of which may be irrelevant to the attack. Likewise, a scientist may test numerous molecules – incurring high cost – while searching for a new drug candidate, many of which may prove useless in a medical setting. In these situations, the goal is to guide

the user to discover as many relevant data points as possible while alleviating information overload caused by irrelevant information.

To identify an algorithm for guiding users in the task of discovery, we examine the family of *active learning* algorithms. Active learning algorithms strategically guide the model during training to fulfil an analytic objective. First, we compare and contrast a set of active learning algorithms and demonstrate that their effectiveness is task-dependent. We demonstrate that choosing an appropriate technique for a given visual analytic task is key to performance. In doing so, we focus our experiments on the VAST challenge 2011 dataset which was designed to mimic a realistic national security scenario [8, 16, 32].

Once we identify active search as the leading active learning algorithm for the task of discovery, we design a visual analytic proof-of-concept tool and augment it with active search. In this prototype, the interactive visualization is the medium of communication between the algorithm and the user. The user inspects data points sequentially and determines their relevance to the task at hand. Simultaneously, the algorithm translates the observed interactions into labels for the underlying models and then guides the user by recommending the most promising data points for investigation.

Although there has been a history of integrating similar ML approaches into visual analytic tools, the impact of these technologies on human behavior and contributing factors to their effectiveness are open for investigation. In a preliminary attempt to address this gap, we conduct a crowd-sourced user study to investigate the impact of the active search algorithm in assisting users during visual data exploration and discovery. Using our proof-of-concept system, participants saw a visualization of geo-tagged microblogs, which provided information about the spread of various disease symptoms. The task was to assist authorities by searching through social media posts to identify individuals who may be impacted by the potential epidemic. We present our user study results, highlighting some of the promises and challenges of this guided data discovery framework. The dataset and analysis scripts are available on GitHub [1].

A summary of our contributions is as follows:

- We present a comparison between three active learning algorithms and demonstrate that their performance is task-dependent. Selecting an appropriate algorithm that maps to a given visual analytic task is critical.
- We map the interactive data discovery task into an active learning algorithm (namely active search) and design an ecosystem where user interactions inform underlying models and the algorithm guides the user through visual clues.
- We present a user study to validate the effectiveness of the active search algorithm in the interactive data discovery workflow. Our results show consistent speedup in discovery throughput and fewer distractions by irrelevant portions of the data.

2 BACKGROUND

In this section, we provide a brief overview on visual analytic themes related to learning from user interactions and guiding users during interactive analysis with a focus on active learning algorithms.

---

*e-mail: monadjemi@wustl.edu
†e-mail: sha@wustl.edu
‡e-mail: alvitta@wustl.edu

---

[1] https://github.com/washuvis/vis2022guideddiscovery

2.1 Active Learning

As datasets grew in size and complexity, the machine learning (ML) community faced the challenge of gathering labeled datasets. With exhaustive labeling no longer being a viable option, ML researchers developed the family of *active learning* algorithms to guide the training process. Active learning traditionally refers to the idea of learning algorithms choosing their training data strategically by querying an oracle in order to achieve better predictive performance [33]. Over time, ML researchers have expanded the active learning paradigm to include new optimization objectives beyond model accuracy. Some examples include the discovery of as many members of a given class as possible (i.e. *active search* [14]) and the detection of anomalous categories (i.e. *rare category detection* [19]). Recently, the visual analytics community has utilized these advances to incorporate humans in the active learning process via interactive interfaces [24, 25, 27]. To the best of our knowledge, this work is the first to incorporate and evaluate the *active search* algorithm in an interactive setting.

2.2 Learning from User Interactions

A significant body of work in the visual analytics community has sought to enable human–machine partnerships in which machines are informed by low-level user interactions with interactive visualizations [1, 3, 5, 11, 29, 31, 36]. Here, the technique of *semantic interaction* is key, where user interactions with visualization tools translate into observations for underlying models, integrating user knowledge in the analysis process and informing intelligent response by the visualization system [13]. Semantic interactions have been used to learn expert knowledge [4], improve visual projection [20], improve text analysis models [24, 25, 34], and mitigate selection bias [15, 28]. For example, Brown et al. [4] proposed Dis-Function, a technique to represent expert knowledge as a distance function which is interactively learned by drag/drop interactions with a 2D visualization. In their work, the user was not guided by an active learning algorithm to decide which points to move. Such decisions fully relied on human judgement. To gauge how human judgement compares with those of active learning algorithms, Bernard et al. [2] designed a study where the goal was to select training data for a classifier. They show that a classifier trained on data selected by humans can outperform one trained on data selected by an active learning algorithm. Their findings rely on the underlying assumption that a 2D visualization of the data is available and separates underlying class distributions. In this work, we evaluate a complete cycle, where the underlying models learn from user interactions and an active learning algorithm guides the user in discovery.

2.3 Providing Visual Analytic Guidance

In addition to using interactions to inform machine learning models, researchers have also developed systems where machines take actions to guide users in the analytic process [6, 12, 17, 18, 22, 30]. Ceneda et al. [7] have defined *guidance* as a computer-assisted process aimed at resolving users' knowledge gap during an interactive session. For example, Kim et al. [24] proposed TopicSifter, an interactive system with the primary purpose of building models with high recall on text documents. This is an example of a system which guides the user by recommending potentially relevant keywords to include in the search space. Another example is ALVA by Kucher et al. [25] which trains a stance classification model by querying the human user for labels. Yet another example of incorporating active learning into visual analytic workflows is RCLens by Lin et al. [27]. Their technique utilizes the rare category detection algorithm [19] to guide users in identifying anomalies. In the context of our work, we utilize an active learning algorithm to guide the user to discover relevant data points efficiently while avoiding irrelevant ones.

3 SELECTING AN ACTIVE LEARNING ALGORITHM

Our primary focus in this section is the *strategy* for selecting data points and their effectiveness in accomplishing a given analytic task. Given a set of unlabeled data points, active learning algorithms sequentially pick one (or more) data points according to a *strategy*. The selected data points are then presented to an oracle to acquire their labels. This oracle may be a human user or an expensive procedure (e.g. laboratory experiments). The observed labels are then used to re-train the underlying models.

In the simplest case, the learning algorithm selects data points from the set of unlabeled data at random. This strategy, known as *random search*, does not aim to optimize any particular objective and is often used as a baseline in ML literature. When the analytic goal is to train an accurate model, one approach is to identify points with the highest uncertainty and acquire their labels. This strategy, known as *uncertainty sampling*, prioritizes data points on the boundary of classes where the model may be most confused about the labels [26]. In some application areas, the primary objective is not to train an accurate model. Instead, the analyst seeks to identify as many members of a valuable class as possible. For such scenarios, Garnett et al. [14] have proposed an active learning approach called *active search*. We note that there are more active learning algorithms available in the ML literature. We focus on these three for demonstration purposes. To show that choosing the appropriate active learning strategy for a given analytic goal is critical, we conduct simulations using the VAST challenge 2011 dataset (described in Sec. 3.1). These experiments do not aim to simulate user behavior, but rather evaluate algorithm behavior for a set of analytic objectives.

3.1 Scenario and Dataset

In VAST Challenge 2011, the fictitious city of Vastapolis is under a biochemical attack, initiating an epidemic with flu-like symptoms. The health officials have access to a large collection of geo-tagged tweet-like microblogs. They want to identify impacted individuals, impacted neighborhoods, and the symptoms from social media posts. In doing so, they need to identify as many illness-related microblogs as possible. We select a sample of 50,000 microblogs where only 1.3% of them contain illness related content. We consider the illness related microblogs to be *relevant* and the remaining ones to be *irrelevant*. We encode the microblogs into a numerical space using an off-the-shelf *word2vec* model, and build a k-NN binary classifier using the cosine similarity metric. This model provides us with a mechanism to reason about the relevance of an unlabeled microblog in light of past observations. We cross validated these design choices against a set of alternatives to ensure our model is a viable choice.

3.2 Simulation Setup

Starting with one illness-related data point as our training set, we make 250 queries according to each of the following active learning strategies/algorithms: random search, greedy active search, and uncertainty sampling. We repeat this experiment 100 times and report evaluation metrics corresponding to three analytic objectives: **Discovery** (the number of illness-related microblogs identified within the 250 queries), **Detection** (the number of unique symptoms detected within the 250 queries), and **Training** (the ROC-AUC score of a k-NN binary classifier trained on the 250 queried data points).

Table 1: Comparing three active learning strategies for three analytic goals. The evaluation metrics are described in Section 3.2.

	Discovery	Detection	Training
	# of microblogs	# of symptoms	ROC-AUC
Random Search	3.3 ± 0.43	3.3 ± 0.43	0.62 ± 0.01
Greedy Active Search	196.2 ± 13.41	19.8 ± 1.23	0.82 ± 0.01
Uncertainty Sampling	63.8 ± 5.00	17.8 ± 1.26	0.84 ± 0.01

71

Figure 1: A view of the prototype system for the epidemic dataset.

3.3 Simulation Results

The outcomes of our simulations are shown in Table 1. It is evident that different active learning strategies are dominant for different analytic objectives. When the objective is to discover members of a valuable class efficiently, the active search strategy outperforms the other strategies significantly. On the other hand, when the objective is to train an accurate classifier, the uncertainty sampling strategy outperforms the other strategies. Our results paint a consistent picture that choosing an appropriate active learning strategy for a task at hand is critical for performance.

4 EXAMPLE PROTOTYPE AND USER STUDY

To investigate the impact of active search guidance on user behavior during visual exploration and discovery, we designed a crowd-sourced user study [2] that tasked participants to interact with a map of the fictional city of Vastapolis (Fig. 1), visualizing microblogs as dots placed on their posting locations. Hovering on data points triggered a tooltip containing the microblog and a bookmark button (Fig. 1, C). A countdown of the remaining time was shown, and users had the option to exit the experiment or report issues (Fig. 1, B). User bookmarks were listed in the side bar (Fig. 1, A).

4.1 Task

Participants were told that health professionals had reported a spike in illnesses with flu-like symptoms, including fever, chills, sweats, nausea and vomiting, and diarrhea. We informed participants that the authorities are interested in identifying the impacted parts of the city by analyzing social media activity. Their task was to assist the authorities by searching through a dataset of geo-tagged microblogs using the interactive interface shown in Figure 1 and bookmarking as many posts containing illness-related information as possible. For this study, we narrowed the dataset to a sample of 3000 points from the approximate start of the epidemic, where 33% of the data points contained illness-related content. We made these design choices to avoid an overly-crowded map and ensure that the discovery task is not prohibitively difficult for the participants.

4.2 Participants

We recruited 130 participants via Amazon's Mechanical Turk platform. Participants were 18 to 65 years old, from the United States, and fluent in English. Each participant had a HIT approval rating of greater than 98% with more than 100 approved HITs. After data cleaning steps outlined in Section 4.4, there were 46 women, 76 men, and 1 participant with undisclosed sex in our subject pool with ages ranging from 18 to 62 years ($\mu = 36$, $\sigma = 9$). 72% of our participants self-reported to have at least an associate degree. The average completion time (including reading the tutorial, performing

---

the task, and completing the survey) was 12 minutes. The instructions specified that participants will be compensated $1.00 base pay and an additional $0.10 bonus for every relevant microblog they identify (with a maximum of $4.00). Although the advertised payment structure was designed to incentivize participants to complete the task, we paid everyone the maximum bonus of $4.00 for fairness.

4.3 Procedure

The experiment complied with an approved protocol per Washington University's IRB. Workers who accepted the HIT followed a URL to the study platform. Our system randomly assigned each participant to one of the following groups: the *active search group,* which received a batch of 10 active search queries in the form of visual clues (orange dots on the map, as shown in Fig. 1) that were updated after every bookmark interaction, and the *control group,* which did not receive any assistance during exploration. Upon giving consent to participate in our study, participants were given a tutorial on their task and their corresponding system. Both groups initiated their task without any initial "clues," and in particular the active search group did not receive assistance for selecting their first bookmark. Participants were given at most 10 minutes to identify as many microblogs related to the epidemic as they could using the interactive tool. Once the users were either satisfied with their search or the 10 minutes were up, they were directed to a post-experiment survey to collect demographic data, feedback on the system, and they self-reported their trust towards the suggestions on a Likert scale. In case our participants experienced technical difficulties, we provided an option to report issues, exit the session, and receive compensation.

4.4 Data Collection, Cleaning, and Exclusions

We analyze our user study data by focusing on two interactions: inspection of microblogs (*hovers*) and discovery of relevant ones (*bookmarks*). These two types of interactions inform us about the speed and accuracy of visual data discovery through the metrics listed in Table 2. Throughout this analysis, we consider illness-related microblogs to be *relevant*. The bookmark and hover purity metrics are the proportion of all bookmarks and all hovers that involved relevant data points, respectively. The bookmarks- and hovers-per-minute metrics inform us about the speed at which users interacted with data points. The relevant hovers and relevant bookmarks-per-minute metrics are the rate at which users interacted with relevant data points, quantifying both speed and accuracy. Finally, we measure the number of relevant microblogs discovered by the end of the session and the number of unique symptoms in the discovered microblogs.

In a pre-processing step, we filtered the collected data to exclude participants who did not attempt the task or were unable to finish the experiment. Specifically, we eliminated participants based on the following four criterion: (1) failed the survey attention checks (1 subject), (2) reported technical issues (1 subject), (3) hovered[3] on fewer than 10 data points (4 subjects), and (4) did not meet the age qualification (1 subject). A total of 123 subjects remained after filtering (74 in the *control* group and 49 in the *active search* group).

4.5 Results

Suggestion Quality: We begin our analysis by examining the quality of suggestions provided by the active search algorithm when seeded with users' interaction data. We define suggestion purity to be the proportion of unique *relevant* microblogs recommended to the user throughout a given session. On average, active search group participants had a suggestion purity of 79%. We observe a moderate positive correlation between bookmark purity and suggestion purity ($R^2_{adj} = 0.594$, $p < 0.0001$), suggesting that the active search algorithm provides useful recommendations for participants who interacted with microblogs containing known symptoms.

---

[3]We consider a valid hover to be one that lasts at least 500ms (300ms for triggering the tooltip, and 200ms for skimming the text)

Table 2: The results of two-sample t-tests on the metrics discussed in Section 4.4

| | 95% CI | | | | |
Metric	Control N=74	Active Search N=40	p-value	t-statistic	Cohen's d
Hovers per Minute	16.7 ± 1.19	14.3 ± 1.23	**0.0112**	-2.58	-0.51
Relevant Hovers per Minute	6.7 ± 0.68	9.2 ± 1.12	**0.0001**	4.00	0.79
Hover Purity	0.39 ± 0.02	0.63 ± 0.05	$<$ **0.0001**	9.70	1.92
Bookmarks per Minute	6.9 ± 0.77	9.5 ± 1.41	**0.0006**	3.52	0.70
Relevant Bookmarks per Minute	5.4 ± 0.68	8.1 ± 1.26	**0.0001**	3.98	0.79
Bookmark Purity	0.77 ± 0.04	0.82 ± 0.05	0.2249	1.22	0.24
Relevant Microblogs Bookmarked	53.9 ± 6.80	73.4 ± 11.50	**0.0026**	3.09	0.61
Unique Keywords Identified	16.1 ± 0.83	15.6 ± 1.32	0.4980	-0.68	-0.13

Suggestion Usage: We observed an unexpected pattern in the active search group. As shown in Figure 2, for approximately 24% of participants in the active search group, suggested microblogs accounted for less than 10% of their bookmarks. 9 out of 49 participants did not bookmark any of the suggestions presented to them at all. Further inspection reveals that the 9 active search participants who ignored the suggestions had on average $82 \pm 9\%$ suggestion purity and $76 \pm 11\%$ bookmark purity. This compares to the the 40 active search participants who interacted with the suggestions, who had on average $79 \pm 5\%$ suggestion purity and $82 \pm 5\%$ bookmark purity. Finally, we observed a difference between how subjects reported their trust towards system suggestions on a 1–5 Likert scale in the post-experiment survey (3.3 ± 0.46 for those who ignored suggestions vs. 4.2 ± 0.24 for those who interacted with the suggestions).
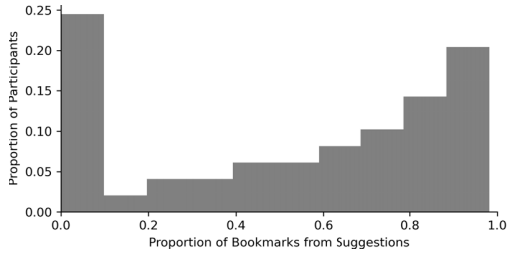

Figure 2: Distribution of proportion of bookmarks resulting from the active search suggestions.

Moving forward, our analyses focus on the impact of suggestions on data exploration and discovery. Thus, we exclude the 9 participants in the *active search* group who did not interact with the suggestions, leaving us with 74 participants in *control* group and 40 participants in the *active search* group. This additional filtering step further cleans our data to compare participants who had access to and used system recommendations to those who did not have access to recommendations. This additional filtering step does *not* impact the conclusivity of our results in Table 2 (see the supplemental material).
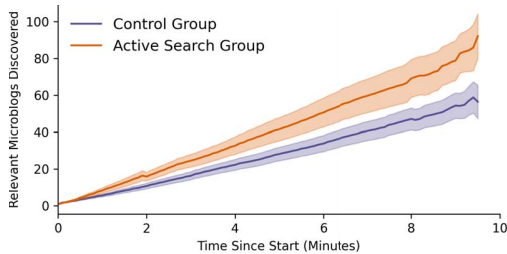

Figure 3: The average number of relevant microblogs discovered and the 95% confidence interval for each group.

The Effect of Suggestions on Data Discovery: We performed a series of two-sample t-tests to investigate differences in behavior in our two study conditions: *control* and *active search*. Table 2 summarizes our findings. We found that participants in the *active search* group bookmarked and hovered over significantly more relevant microblogs per minute than the *control* group. Furthermore, our findings show that the *active search* group performed fewer exploratory hovers per minute than the *control* group, implying that the suggestions resulted in a more efficient exploratory analysis. For a more fine-grained analysis, we examine bookmark discoveries as a function of time. Figure 3 shows the average number of bookmarks over time for the *active search* and *control* groups. We observe that the *active search* group consistently outperformed the control group by bookmarking more relevant microblogs throughout the ten-minute session. However, it is noteworthy that although the suggestions improved the quantity of the bookmarks, we found no measurable difference in the quality or content of the bookmarked discoveries. Both the *active search* and *control* groups collectively examined similar geographical regions and symptom sets.

Impact of Active Search Suggestions on Usability: In a post-experiment survey, we asked subjects in both groups three questions on *willingness to use*, *ease of use*, and *ease of task completion*. We performed a Mann–Whitney U statistical test at $\alpha = 0.05$ to determine if there was a significant difference between the control and active search groups. The analysis showed some evidence that the active search group found the system easier to use ($U = 1199.50$, $p = 0.0336$, $r = 0.16$) and were more willing to use the system frequently ($U = 1192$, $p = 0.0362$, $r = 0.16$). However, the effect sizes were small for both. Furthermore, we did not find a significant difference between the control and active group's response to ease of task completion ($U = 1397.50$, $p = 0.2989$, $r = 0.07$).

5 DISCUSSION AND CONCLUSIONS

We considered the task of discovery: identifying relevant data points to the task at hand efficiently. Upon examining the family of active learning algorithms, we selected active search as the leading technique for guiding this discovery process. Our user study results validate that guidance provided by the active search algorithm can significantly improve interactive data discovery. Our findings show that the participants successfully disregarded *irrelevant* information and were more mindful towards the *relevant* data points. These results can have high-impact implications for designing visual analytic tools for tasks such as intelligence analysis and scientific discovery. Unexpectedly, we observed a nontrivial number of participants who ignored active search recommendations despite their relevance to their task. This highlights an avenue for investigating how to present such recommendations effectively in similar visual analytic systems.

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*, pp. 1363–1375, 2016.

[2] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE transactions on visualization and computer graphics*, 24(1):298–308, 2017.

[3] Y. Bian, J. Wenskovitch, and C. North. DeepVA: Bridging Cognition and Computation through Semantic Interaction and Deep Learning. *Machine Learning from User Interactions for Visualization and Analytics at IEEE VIS*, 2019.

[4] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-Function: Learning Distance Functions Interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 83–92. IEEE Computer Society, 2012.

[5] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang. Finding Waldo: Learning about Users from their Interactions. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1663–1672, 2014.

[6] D. Ceneda, N. Andrienko, G. Andrienko, T. Gschwandtner, S. Miksch, N. Piccolotto, T. Schreck, M. Streit, J. Suschnigg, and C. Tominski. Guide me in analysis: A framework for guidance designers. In *Computer Graphics Forum*, vol. 39, pp. 269–288. Wiley Online Library, 2020.

[7] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, and C. Tominski. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, 2016.

[8] K. Cook, G. Grinstein, and M. Whiting. The vast challenge: History, scope, and outcomes: An introduction to the special issue, 2014.

[9] R. J. Crouser and R. Chang. An Affordance-Based Framework for Human Computation and Human-Computer Collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2859–2868, 2012.

[10] R. J. Crouser, A. Ottley, and R. Chang. Balancing Human and Machine Contributions in Human Computation Systems. In *Handbook of Human Computation*, pp. 615–623. Springer, 2013.

[11] 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.

[12] S. Das, D. Cashman, R. Chang, and A. Endert. Gaggle: Visual analytics for model space navigation. 2019.

[13] A. Endert, P. Fiaux, and C. North. Semantic Interaction for Visual Text Analytics. In *Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems*, pp. 473–482, 2012.

[14] R. Garnett, Y. Krishnamurthy, X. Xiong, J. Schneider, and R. Mann. Bayesian Optimal Active Search and Surveying. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[15] D. Gotz, S. Sun, and N. Cao. Adaptive Contextualization: Combating Bias During High-Dimensional Visualization and Data Selection. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 85–95, 2016.

[16] G. Grinstein, K. Cook, P. Havig, K. Liggett, B. Nebesh, M. Whiting, K. Whitley, and S. Knoecni. Vast 2011 challenge: Cyber security and epidemic. *IEEE VAST*, 2011:299–301, 2011.

[17] S. Guo, F. Du, S. Malik, E. Koh, S. Kim, Z. Liu, D. Kim, H. Zha, and N. Cao. Visualizing uncertainty and alternatives in event sequence predictions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[18] S. Guo, Z. Jin, Q. Chen, D. Gotz, H. Zha, and N. Cao. Visual anomaly detection in event sequence data. In *2019 IEEE International Conference on Big Data (Big Data)*, pp. 1125–1130. IEEE, 2019.

[19] J. He and J. Carbonell. Nearest-neighbor-based active learning for rare category detection. In *21st Annual Conference on Neural Information Processing Systems, NIPS 2007*, 2009.

[20] T. Iwata, N. Houlsby, and Z. Ghahramani. Active Learning for Interactive Visualization. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, pp. 342–350, 2013.

[21] S. Jiang, G. Malkomes, G. Converse, A. Shofner, B. Moseley, and R. Garnett. Efficient Nonmyopic Active Search. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 1714–1723, 2017.

[22] Z. Jin, S. Cui, S. Guo, D. Gotz, J. Sun, and N. Cao. Carepre: An intelligent clinical decision assistance system. *ACM Transactions on Computing for Healthcare*, 1(1):1–20, 2020.

[23] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual Analytics: Definition, Process, and Challenges. In *Information Visualization*, pp. 154–175. Springer, 2008.

[24] H. Kim, D. Choi, B. Drake, A. Endert, and H. Park. TopicSifter: Interactive Search Space Reduction through Targeted Topic Modeling. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 35–45. IEEE Computer Society, 2019.

[25] K. Kucher, C. Paradis, M. Sahlgren, and A. Kerren. Active Learning and Visual Analytics for Stance Classification with ALVA. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3):1–31, 2017.

[26] D. D. Lewis and J. Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994*, pp. 148–156. Elsevier, 1994.

[27] H. Lin, S. Gao, D. Gotz, F. Du, J. He, and N. Cao. Rclens: Interactive rare category exploration and identification. *IEEE transactions on visualization and computer graphics*, 24(7):2223–2237, 2017.

[28] S. Monadjemi, R. Garnett, and A. Ottley. Competing Models: Inferring Exploration Patterns and Information Relevance via Bayesian Model Selection. In *2020 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE Computer Society, 2020.

[29] 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.

[30] I. Pérez-Messina, D. Ceneda, M. El-Assady, S. Miksch, and F. Sperrle. A typology of guidance tasks in mixed-initiative visual analytics environments. 2022.

[31] 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.

[32] J. Scholtz, M. A. Whiting, C. Plaisant, and G. Grinstein. A reflection on seven years of the vast challenge. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualization*, pp. 1–8, 2012.

[33] B. Settles. Active Learning Literature Survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.

[34] R. Sevastjanova, M. El-Assady, A. Hautli-Janisz, A.-L. Kalouli, R. Kehlbeck, O. Deussen, D. A. Keim, and M. Butt. Mixed-Initiative Active Learning for Generating Linguistic Insights in Question Classification. *Data Systems for Interactive Analysis at IEEE VIS*, 2018.

[35] M. A. Siddiqui, A. Fern, T. G. Dietterich, R. Wright, A. Theriault, and D. W. Archer. Feedback-Guided Anomaly Discovery via Online Optimization. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2200–2209, 2018.

[36] K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch. Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum*, 39(3):757–783, 2020.